

Univerzita Karlova v Praze
Přírodovědecká fakulta

Studijní program: Buněčná a vývojová biologie
Studijní obor: Vývojová biologie



Mgr. Eva Seifertová

Genetické mapování u rodu *Xenopus*
Genetic mapping in *Xenopus*

Dizertační práce

Vedoucí závěrečné práce/Školitel: RNDr. Ing. Vladimír Krylov, Ph.D.

Praha, 2014

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval/a samostatně a že jsem uvedl/a všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 28. 2. 2014

Podpis

Abstrakt

Mezi nejvýznamnější modelové organizmy v oblasti vývojové biologie patří bezesporu obojživelník *Xenopus tropicalis*. Jeho dřívější využití především v embryologickém výzkumu je v současnosti vytlačováno studii genetického a genomického charakteru. *X. tropicalis* má deset párů chromozómů v diploidním genomu a proto je pro tento typ výzkumu velmi vhodný. V posledních deseti letech byl jeho genom osekvenován, několikrát sestaven, vznikla provizorní genetická mapa a byla vytvořena i BAC knihovna mnohonásobně pokrývající genom. Přes veškeré úsilí se u tohoto druhu doposud nepodařilo sestavit kompletní mapu genomu. Ten i nadále zůstává organizován ve formě scaffoldů, které mají často neznámou polohu a i jejich sestavení zůstává i nadále sporné.

Náš výzkum byl zaměřen na kompletaci genomu u druhu *Xenopus tropicalis* a na nové přístupy, které by bylo možné využít i u jiných druhů. Nejprve byla na základě vazebné analýzy a zjištěné fyzické polohy markerů sestavena genetická mapa. Ta nebyla zcela kompletní- nezahrnovala krátké raménko chromozómu 2 a rovněž 15 cM z p raménka chromozómu 7. Protože bylo zaplnění těchto oblastí klasickými metodami velmi obtížné, nebo ještě pravděpodobněji zcela nemožné, byla vynalezena nová metoda pro genetické mapování. Ta zahrnuje mikrodisekci zvolené oblasti, celogenomovou amplifikaci získané DNA a její osekvenování metodou Illumina. Postup mikrodisekce a amplifikace byl využit ve srovnávací studii příbuzných druhů *X. tropicalis* a *X. laevis* metodou *Zoo-FISH*, kdy amplifikovaná DNA z druhu *X. tropicalis* sloužila jako sonda a ukázala sekvenční podobnost meiotických kvartetů.

Ve třetí studii bylo mikrodisektováno, amplifikováno a sekvenováno 15 kopií krátkého raménka chr.7. Získaná čtení byla srovnána se současnými verzemi genomu *X. tropicalis*. K vyříznuté oblasti se nejen podařilo přiřadit velké množství sekvencí a genů, ale byly rozpoznány i špatně sestavené scaffoldy. Data získaná z tohoto výzkumu byla použita i pro studii srovnávající rozdíly pohlavních chromozómů u tohoto druhu. Mikrodisekce s následnou sekvenací byla pro účely genetického mapování použita vůbec poprvé a je zřejmé, že je tento přístup použitelný i u jiných živočišných druhů, a to jak s osekvenovaným, tak s neznámým genomem. Metodiku je ovšem možné použít i v jiných oblastech, jako je klinická medicína či onkologie.

Abstract

The diploid amphibian *Xenopus tropicalis* represents a significant model organism for studies of early development, genes function and evolution. Such techniques as gynogenesis, injection of morpholino antisense oligonucleotide into fertilized eggs or transgenesis were established. In the recent ten years, many efforts have been made to complete the sequence information. *X. tropicalis* genome has been sequenced but the completion of its assembly only on the basis of sequence data has been impossible. Therefore, our first work was focused on one of approaches for a genome completing- genetic mapping. First of all, the genetic map of *Xenopus tropicalis* was established pursuant linkage and physical positions of markers. Since the map contained gaps, we developed a new method for genetic mapping based on the next generation sequencing of laser microdissected arm. Using Illumina next generation sequencing of fifteen copies of a short arm of chromosome 7, we obtained new insights into its genome by localizing previously unmapped genes and scaffolds as well as recognizing mislocalized portions of the genome assembly. This was the first time laser microdissection and sequencing of specific chromosomal regions has been used for the purpose of genome mapping. These data were also used in the evolution study of the sex determining area placed on the q arm of chromosome 7, which showed that *Xenopus tropicalis* sex chromosomes contain large pseudoautosomal areas. Moreover, we made Zoo-FISH analysis using *X. tropicalis* microdissected chromosomes as probes for labeling *Xenopus laevis* chromosomes, which revealed similarity of meiotic quartets even after 65 million years of separate evolution.

Our novel approach for next generation sequencing of microdissected chromosomal area is also applicable to species without sequenced genomes or for clinical applications in medical cytogenetics and oncology where tissue availability may be limiting. This method is likely to be of widespread use in species where individual chromosomes are distinguishable by cytological methods.

Obsah

Abstrakt.....	3
Abstract.....	4
Seznam zkratk	7
1. Literární úvod.....	8
1.1 Úvod	8
1.2. Taxonomie.....	8
1.2.1 <i>Xenopus tropicalis</i>	9
1.2.1.1 Používané linie.....	9
1.2.1.2 Názvosloví chromozómů	9
1.2.1.3 Určení pohlaví.....	10
1.2.2 <i>X. laevis</i>	10
1.2.2.1 Chromozómy.....	10
1.2.2.2 Určení pohlaví.....	11
1.3 Genomika	11
1.3.1 Genetická mapa.....	11
1.3.1.2 Vazebné mapování	11
1.3.1.2.1 Vazebná mapa u <i>Xenopus laevis</i>	12
1.3.1.2.2 Vazebná mapa u <i>Xenopus tropicalis</i>	13
1.3.1.3 Fyzická mapa	13
1.3.1.3.1 FISH sondy	14
1.3.1.4 Zoo-FISH	14
1.3.2 Sekvenování	15
1.3.2.1 Chemické sekvenování (Maxam-Gilbert sequencing)	15
1.3.2.2 Řetězcová metoda (Sanger sequencing).....	15
1.3.2.3 Sekvenování „nové generace“ (Next Generation Sequencing, NGS)	16
1.3.2.3.1 MPSS (Massively Parallel Signature Sequencing)	16
1.3.2.3.2 454- pyrosekvenování	17
1.3.2.3.3 Illumina (dříve Solexa) sekvenování.....	17
1.3.3 Kompletace genomu (Assembly)	17
1.3.3.1 Genom <i>Xenopus tropicalis</i>	19
1.3.3.1.1 Verze 4.1 (v.4.1)	19
1.3.3.1.2 Verze 7.1 (v.7.1)	19
1.3.3.2 Genom <i>Xenopus laevis</i>	20

1.3.3.3 Zdroje informací	20
1.4 Separace chromozómů.....	20
1.4.1 Mikrodisekce.....	21
1.4.1.2 Laserová mikrodisekce.....	21
1.4.2 Průtoková cytometrie.....	21
1.5 Celogenomové amplifikace (Whole Genome Amplification, WGA)	22
1.5.1 PEP (Primer Extension Preamplification).....	22
1.5.2 DOP-PCR (Degenerate Oligonucleotide Primed- Polymerase Chain Reaction)	23
1.5.3 MDA (Multiple Displacement Amplification)	23
1.5.4 OmniPlex WGA.....	24
1.5.5 Nevýhody celogenomových amplifikací a jejich srovnání	24
1.6 Bioinformatika	24
1.6.1 Algoritmy pro vyhodnocování dat ze sekvenací.....	24
1.6.2 Datové formáty pro zaznamenání sekvence.....	25
1.6.2.1 FASTA	25
1.6.2.2 FATSQ	26
1.6.2.3 Další formáty databází	26
1.6.2.3.1 BED formát.....	26
1.6.3 Základní programovací jazyky používané v bioinformatice.....	27
1.6.3.1 PERL	27
1.6.3.2 PYTHON.....	27
2. Cíle práce.....	28
3. Výsledky	29
4.1 Genetická mapa.....	66
4.2 Příprava malovacích sond druhu <i>X. tropicalis</i> z mikrodisektovaných chromozómů a rekonstrukce tetraploidního karyotypu <i>X. laevis</i> metodou Zoo-FISH.....	68
4.3 Efektivní vysokorychlostní sekvenace laserově mikrodisektovaného chromozomálního ramene.....	70
4.4 Velké pseudoautosomální oblasti na pohlavních chromozómech žáby <i>Silurana (Xenopus) tropicalis</i>	72
5. Závěr	74
6. Použitá literatura	76

Seznam zkratek

AFLP	Amlified Fragment Lenght Polymorphism
BAC	Bacterial Arteficial Chromosome
cM	centimorgan
DOP-PCR	Degenerate Oligonucleotide Primed- Polymeraze Chain Reaction
FACS	Fluorescent Activated Flow Cytometry, fluorescenčně aktivovaná průtoková cytometrie
FISH	Fluororesent In Situ Hybridization, fluorescenční in situ hybridizace
FISH-TSA	FISH coupled with Tyramide Signal Amplification, fluorescenční in situ hybridizace spojená s tyramidovou amplifikací
I-PEP	Improved- PEP
IR	Infračervené záření
ISH	In Situ Hybridizace
LG	Linkage Group, vazebná skupina
LP-DOP-PCR	Long Products DOP-PCR
MDA	Multiple Displacement Amplification
MPSS	Massively Parallel Signature Sequencing
NGS	Next Generation Sequencing, sekvenování nové generace
PAC	P1- derived arteficial chromosome
PEP	Primer Extension Preamplification
RCD	Relative Distance From Centromere, relativní vzdálenost od centromery
SSLP	Simple Sequence Length Polymorphysm
UV	ultrafialové záření
v.4.1	Kompletace genomu <i>Xenopus tropicalis</i> verze 4.1
v.7.1	Kompletace genomu <i>Xenopus tropicalis</i> verze 7.1
WGA	Whole genome amplification, celogenomová amplifikace
WGA3	GenomePlex DNA Reamplifying Kit
WGA4	GenomePlex Single cell WGA Kit
Xl (1-18)	Chromozóm druhu <i>X. laevis</i> s příslušným číslem
Xt (1-10)	Chromozóm druhu <i>X. tropicalis</i> s příslušným číslem
YAC	Yeast Arteficial Chromosome
Zoo-FISH	mezidruhová Fluoroesenční In Situ Hybridizace

1. Literární úvod

1.1 Úvod

V posledních deseti až patnácti letech se *Xenopus tropicalis* dostal do popředí zájmu nejen buněčných a vývojových biologů, ale také bioinformatiků. Jeho genom byl osekvenován, bylo publikováno několik verzí assembly a rovněž byla vytvořena genetická mapa. V současné době je známo a osekvenováno přibližně 50% z celkového množství cDNA a asi 35% genů v jeho genomu.

X. tropicalis je rovněž výborným modelovým organizmem pro studování raného vývoje u obratlovců. Je velmi dobře dostupný, produkuje značné množství oocytů a embrya s vnějším vývojem jsou dostatečně velká pro snadnou manipulaci. Z tohoto důvodu byla od sedmdesátých let xenopí embrya a oocyty důležitým modelem pro výzkum molekulárních podstat formace os, indukce embryonálního vývoje, signálních drah v organogenezi a buněčného cyklu.

Velkou výhodou xenopího modelu je rovněž jeho schopnost produkovat proteiny už v raném stádiu vývoje. To bylo využíváno při studiích funkcí jednotlivých proteinů. Po mikroinjikace libovolné mRNA bylo možné sledovat over- expresi příslušného genu a přímé následky tohoto procesu (Smith and Harland 1991). Podobnou, ovšem opačnou metodou je i injikace morpholina- antisense oligonukleotidu, který příslušnou mRNA naopak umlčí (Scacheri et al. 2004). Tento systém je možné využít i pro umlčení několika genů v jednom embryu (Khokha and Loots 2005) a u druhu *X. tropicalis* se jedná o nejoblíbenější a nejvyužívanější způsob umlčování genů.

Nezbytnou součástí takových funkčních studií je samozřejmě sekvenční znalost příslušné mRNA. V několika posledních letech zaznamenala xenopí komunita významné pokroky v různých oblastech genetiky. Jedním z nejvýznamnějších počínů bylo sestavení meiotické mapy (Wells et al. 2011) a uspořádání jeho osekvenovaného genomu (assembly v.4.1) (Hellsten et al. 2010). Za zmínku ovšem stojí i první souhrnné analýzy indukovaných mutací u obojživelníků a klonování nových mutací (Abu-Daya et al. 2009), či vytvoření center skladujících xenopí cDNA, protilátky, fosmidy či BAC klony, a to jak v Americe, tak v Evropě.

1.2. Taxonomie

Taxonomie druhu *Xenopus tropicalis* a jeho příbuzných druhů z rodu *Xenopus* je značně kontroverzní. Fylogenetickým výzkumem bylo zjištěno, že jsou *Xenopus tropicalis* a *Xenopus*

epitropicalis příbuznější se druhy *Hymenochirus* (česky drápatečka) a *Pipa* (česky pipa americká), než s ostatními druhy rodu *Xenopus* (Poinar and Cannatella 1988). *X. tropicalis* sice patří do rodu *Xenopus*, ovšem do podrodu *Silurana*. Správný název by měl podle tohoto výzkumu znít *Silurana* (a nikoli *Xenopus*) *tropicalis*. Výzkumy založené na sekvenčním srovnání ribosomální DNA ovšem zase ukázaly bližší příbuznost podrodu *Silurana* s rodem *Xenopus* (de Sá and Hillis 1990). Ve své práci nicméně i nadále používám jméno *Xenopus tropicalis*, a to jak z historických, tak praktických důvodů. Český ekvivalent jména je Drápatka tropická.

Xenopus tropicalis je v posledních deseti letech ve výzkumu využíván více, než jeho příbuzný a dříve velmi oblíbený modelový organizmus *Xenopus laevis*. Hlavním důvodem je především jeho menší a pouze diploidní genom- *X. tropicalis* má 20 chromozómů (2n) oproti 36 (2n) chromozómům u *X. laevis* ($1,7 \times 10^9$ bp oproti $3,1 \times 10^9$ bp). Situace u *X. laevis* je navíc složitější ještě o fakt, že se jedná o alotetraploidní druh (viz níže). Tato vlastnost prakticky vyloučila možnost použití *X. laevis* jakožto modelového organismu pro genetické analýzy. Značnou výhodou *X. tropicalis* je i jeho menší rozměr, nižší generační interval, větší počet oocytů a rovněž je možné využít většinu experimentálních systémů vyvinutých pro *X. laevis*.

1.2.1 *Xenopus tropicalis*

1.2.1.1 Používané linie

V současné době existuje pět různých linií *X. tropicalis*. Pro genetický výzkum jsou významné především dvě linie- Nigerian a Ivory coast. Jak napovídá jejich název, jména linií jsou shodná s oblastí, kde byli odchyceni původní zakladatelé těchto linií. Tyto dvě linie byly použity pro vytvoření vazebné mapy, a to z důvodu jejich malé genetické příbuznosti a tedy i snadného hledání polyploidních markerů.

1.2.1.2 Názvosloví chromozómů

X. tropicalis má 10 párů chromozómů. Původní názvosloví ze sedmdesátých let bylo založeno na poněkud nekonvenčním srovnání délky p a q ramen (Tymowska 1973). Toto názvosloví bylo v roce 2010 změněno na tradičnější, kdy jsou chromozómy označeny podle velikosti- největší 1, nejmenší má v tomto případě 10 (Khokha et al. 2009). Pouze podobně velkým chromozómům 7 a 8 byl přiřazen název na základě poměru p/q ramen, kdy nižší číslo

dostal více metacentrický chromozóm. V literatuře je nicméně možné setkat se s oběma typy názvosloví. V této práci je používána novější terminologie.

(Tymowska 1973)	1	2	9	3	4	8	5	10	6	7
(Khokha et al. 2009)	1	2	3	4	5	6	7	8	9	10

Tabulka 1: Srovnání názvosloví chromozómů podle (Tymowska 1973) a (Khokha et al. 2009). 1-10: názvy chromozómů

1.2.1.3 Určení pohlaví

X. tropicalis má pohlaví určené geneticky. Dlouhou dobu nebylo zřejmé, je-li toto určení typu savčího či ptačího (XX/XY nebo ZW/ZZ). Stejně jako další druhy obojživelníků má tento druh nerozlišitelné (homomorfní) pohlavní chromozómy. Pravděpodobnou příčinou rozšíření homomorfních chromozómů je evolučně častá translokace v sex určující oblasti (Schmid et al. 2010, Evans and Charlesworth 2013). Rovněž je možné, že pohlavní chromozómy u těchto druhů mohou rekombinovat ve většině své délky (Stöck et al. 2011). Takové pohlavní chromozómy pak mají velké pseudoautozomální oblasti a jejich rekombinace zabraňuje vzniku pohlavně odlišných chromozómů.

Jediný známý gen určující pohlaví u obojživelníků je DM-W. *X. tropicalis* nicméně tento gen postrádá. Přestože byl osekvenovaný genom *X. tropicalis* samičí (Hellsten et al. 2010), nebyla charakterizována oblast určující pohlaví. Ta byla určena až prokázáním vazby 22 AFLP (Amlified Fragment Lenght Polymorphism) markerů na samičí pohlaví (Olmstead et al. 2010). 4 z těchto markerů se 100% vazbou na pohlaví byly umístěny na distálním konci chromozómu 7 (Wells et al. 2011). Mnoho markerů s vazbou na pohlaví se nicméně nacházelo na malých scaffoldech, které nebyly součástí vazebné mapy a jejich poloha byla tudíž neznámá.

1.2.2 *X. laevis*

1.2.2.1 Chromozómy

Xenopus laevis má 36 chromozómů (2n). Většina z nich může být uspořádána do kvartetů podle podobných replikačních profilů (Schmid and Steinlein 1991). Z toho by bylo možné usuzovat, že je *X. laevis* tetraploidní. Na druhou stranu byla ovšem zjištěna nezávislá dědičnost duplikovaných genů- funkčně je tedy *X. laevis* z větší části diploidizován (Graf and Kobel 1991).

Tento druh pravděpodobně vznikl křížením, ve kterém byly zachovány oba dva rodičovské genomy. Cytologicky je *X. laevis* popsán jako funkční diploid alotetraploidního původu s 36 chromozomy a 18 bivalenty. Jeho genom je velký přibližně $3,1 \times 10^9$ bp (Thiébaud and Fischberg 1977), což je přibližně dvojnásobná velikost oproti genomu *X. tropicalis*.

1.2.2.2 Určení pohlaví

Stejně jako *X. tropicalis* má i *X. laevis* pohlaví určeno geneticky se samičím heteromorfním pohlavím (Mikamo and Witschi 1963). Genem určujícím pohlavím je DM-W, který má 89% identitu s genem DMRT určujícím pohlaví u ptáků či u medaky (Yoshimoto et al. 2006). Tento gen vznikl pravděpodobně poté, co se podrod *Xenopus* oddělil od poddruhu *Silurana*, ale ještě předtím, než došlo k rozlišení druhů *X. laevis* a *X. clivii*. Tento gen byl také nalezen u oktaploidních druhů *X. itombensis* a *X. vestitus* (Bewick et al. 2011).

1.3 Genomika

1.3.1 Genetická mapa

Principiálně existují dva způsoby, kterými je možné vytvořit genetickou mapu. Prvním způsobem je tvořena vazebná mapa a s relativní pozicí dvou či více markerů vytvořená na základě frekvence rekombinace. Druhým přístupem je fyzické mapování, ve kterém je zjišťována skutečná poloha markerů na chromozómech.

1.3.1.2 Vazebné mapování

Jak již bylo řečeno, vazebná mapa je charakterizována frekvencí crossing-overu u homologních chromozómů. 1% četnost rekombinace odpovídá vazebné vzdálenost 1 centimorgan (cM). Je-li vzdálenost mezi dvěma lokusy větší než 50cM, jsou děděny nezávisle a neexistuje mezi nimi vazba. Takové markery se buď nachází na dvou chromozómech, nebo na jednom, ovšem velmi daleko od sebe. Naopak lokusy vazebně velmi blízko u sebe budou téměř vždy děděny společně, protože mají velmi malou pravděpodobnost rekombinace.

Pro tvorbu vazebné mapy jsou používány polymorfní sekvence DNA, které se u jedinců v populaci vyskytují v heterozygotním stavu. Původní genetické mapy byly tvořeny na základě fenotypových projevů ještě před objevením DNA, ovšem vzhledem k velkému množství nekódujících sekvencí v eukaryotickém genomu byly tyto mapy velmi zevrubné, s velkými

vzdálenostmi mezi jednotlivými markery. V současné době je jako DNA markery možné použít prakticky jakékoli heterozygotní lokusy.

Značnou výhodou vazebného mapování oproti fyzickému mapování je možnost namapovat velké množství markerů najednou. Vytvořená mapa může být velmi podrobná. Značnou nevýhodou je ovšem fakt, že mezi vzdáleností v bázích a cM neexistuje lineární vztah. Pravděpodobnost rekombinace mezi dvěma markery závisí i na jejich fyzické poloze na chromozómu, například v okolí centromery je četnost rekombinací velmi malá. Vazebně se tedy mohou dva markery nacházet velmi blízko u sebe, zatímco fyzická vzdálenost je velká. Dalšími místy se sníženou rekombinací mohou být pohlavní chromozómy- vazebné mapy vytvořené zvlášť u samců a samic jsou u lokusů určujících pohlaví u heterogametického pohlaví významně kratší (Kondo et al. 2001). Na pseudoautozomálních oblastech pohlavních chromozómů je naopak pravděpodobnost rekombinace vyšší, celkově jsou tedy mapy nerozlišitelných pohlavních chromozómů stejně dlouhé. U druhu *Medaka* byl tento fakt použit k nalezení pohlavních chromozómů (Kondo et al. 2001).

Při tvorbě vazebné mapy mohou být problematické i ty markery, které se nacházejí na dvou či více chromozómech a přesto jsou děděny společně. Takovému jevu se říká vazebná nerovnováha a může být způsoben mnoha faktory od genetické vazby, selekce, úrovně rekombinace a mutací, genetickým driftem či strukturou populace. Posledním problémem při tvoření vazebné mapy může být samotný typ zvolených markerů, respektive jejich nerovnoměrné rozmístění v genomu. V hraničním případě může dokonce dojít k nezahrnutí některých oblastí do vazebné mapy z důvodu nepřítomnosti zvolených markerů.

1.3.1.2.1 Vazebná mapa u *Xenopus laevis*

Vazebné mapování je u *X. laevis* vzhledem k jeho dlouhé generační době a alotetraploiditě velmi složité. Vazebná mapa tohoto druhu dodnes prakticky neexistuje, ovšem vzhledem k významnému postavení tohoto obojživelníka existovala v minulosti snaha o její vytvoření. První vazebná mapa obsahovala 29 lokusů na celkem 8 vazebných skupinách (LG, Linkage Group) (Graf 1989). Tato mapa byla rozšířena ještě o dvě LG (Graf and Kobel 1991), ovšem vzhledem k celkovému počtu chromozómů (18 párů) je zřejmé, že je tento počet nedostatečný a i pokrytí lokusy je velmi hrubé.

1.3.1.2.2 Vazebná mapa u *Xenopus tropicalis*

X. tropicalis se v posledním desetiletí stal velmi významným druhem právě v oblasti genetiky a proto je u něj situace v porovnání s *X. laevis* zcela odlišná. První vazebná mapa představovala celkem 53 markerů na 13 vazebných skupinách (Kochan et al. 2003). Počet chromozómů tohoto druhu (10 párů) naznačoval, že některé markery segregovaly nezávisle, ačkoli by měly být ve vazbě.

Nejproblematictější oblastí pro vazebné mapování je centromera. U *X. tropicalis* se podařilo vazebně mapovat všech deset centromerických oblastí (Khokha et al. 2009). Bylo k tomu využito gynogenetické křížení, při kterém je zabráněno druhému meiotickému dělení sekundárního oocyty. Vzniklé embryo je tedy diploidní, ale obsahuje pouze mateřské geny. Markery nacházející se blízko centromery jsou ve velkém procentu homozygotní (Hirsch et al. 2002).

Nově vytvořená mapa u druhu *Xenopus tropicalis* (Wells et al. 2011) je součástí této práce a věnuje se jí kapitola 4.1 Genetická mapa.

1.3.1.3 Fyzická mapa

Druhý typ mapy zobrazuje fyzickou polohu genu nebo nějaké jiné DNA sekvence na chromozomu. V dnešní době je nejvyužívanější technikou mapování pomocí FISH sond (Fluorescent In Situ Hybridization). Prvním způsobem fyzického mapování bylo ovšem restriční mapování (Griffin et al. 1974). Nejjednodušším způsobem použití této metody bylo štěpení dvěma restričními endonukleázami- v jedné sérii pokusů zvlášť, v druhé oběma dohromady. Restriční fragmenty byly následně rozděleny podle velikosti na agarózovém gelu a restriční mapa vznikla seřazením jednotlivých fragmentů.

Jak již bylo řečeno, v cytogenetickém výzkumu posledních let se využívá především fluorescenčních sond, tedy označeného vlákna cDNA, které hybridizuje s komplementární sekvencí DNA nebo RNA. Původní technika- In Situ Hybridizace (ISH) (Gall and Pardue 1969) využívala sondy značené jinými způsoby- radiačně, koloidním zlatem nebo chromogenními značkami spojenými s enzymy.

Značnou výhodou ISH a FISH technik je nasedání sondy na cílovou sekvenci nacházející se přímo v buňkách nebo tkáních. Je tak možné zobrazit skutečnou fyzickou polohu hledaného markeru. Na meiotickém či mitotickém profázním jádře je limitní rozlišení metody několik megabází, na interfázích chromozómech asi 100Kbp (Jiang and Gill 2006).

Naopak nevýhodou této techniky je její pracnost- není možné zobrazit víc než několik markerů zároveň a i rozlišení metody je zvláště v případě profázního jádra oproti vazebnému mapování poměrně nízké.

Metody fyzického mapování jsou zcela určitě vhodným doplňkem pro vytvoření genetické mapy, a to pro opravu chyb. Přestože jsou přístupy vazebného a fyzického mapování založeny na zcela jiných principech, pořadí lokusů by mělo být v obou případech stejné. Pokud tomu tak není, musí být chyba ve vazebné mapě nebo v sekvenci.

1.3.1.3.1 FISH sondy

FISH sondy mohou značit RNA či DNA, pro účely mapování se však využívá sond značících DNA přímo v jádře. Sondy mohou být různě specifické a mohou značit jak celý chromozóm, tak i pouze jeden gen. Jako sondy mohou sloužit i kosmidové, BAC (Bacterial Artificial Chromosome), PAC (P1- Derived Artificial Chromosome) či YAC (Yeast Artificial Chromosome) klony. Výhodou mapování těchto větších klonů je možnost namapování velkého množství genů zároveň, ovšem je samozřejmě nutné znát jejich sekvenci. Sekvenace a sestavení genomu klonu o velikosti 200kb je ovšem oproti sekvenaci a následné assembly celého genomu podstatně jednodušší. Nevýhodou velkých sond je častý obsah repetitivních sekvencí, které zvyšují nespecifičnost signálu. Problém je možné vyřešit inkubací sondy s frakcí DNA obohacenou o repetitivní sekvence, případně s neznačenou kompetitorovou DNA studovaného organismu.

1.3.1.4 Zoo-FISH

Zoo FISH je technika pro srovnávací fyzické mapování. Jejím principem je použití DNA sond pocházejících z jednoho živočišného druhu na jiný druh (Scherthan et al. 1994). Většinou se používá u příbuzných druhů, protože vysoké zastoupení homologních sekvencí značně usnadňuje její použití. Nejčastěji jsou používány celochromozómové sondy, případně sondy značící nějakou chromozomální část. Metoda Zoo-FISH je používána především pro evoluční studie srovnávající velké oblasti, protože genomickými algoritmy není doposud možné takové studie udělat. Značnou výhodou této techniky je i skutečnost, že není nutná znalost sekvence studovaných organismů.

1.3.2 Sekvenování

Sekvenování je postup, kterým je určeno pořadí a typ sekvencí v příslušné DNA molekule. Původní metody ze sedmdesátých let byly v posledním desetiletí nahrazeny modernějšími postupy, tzv. sekvenováním nové generace (Next Generation Sequencing, NGS), které jsou výrazně rychlejší a levnější. Nevýhodou některých typů NGS je ovšem výrazně kratší čtení (100-200bp), a u některých velká chybovost (až 15%). Z tohoto důvodu se i nadále využívá sekvenování „první generace“ typu Sanger (až 2000bp s méně než 2% chyb).

1.3.2.1 Chemické sekvenování (Maxam-Gilbert sequencing)

Nejstarší metoda sekvenace byla vynalezena na přelomu let 1976 a 1977 (Maxam and Gilbert 1977). Její princip spočívá v chemickém štěpení řetězce DNA na specifickém místě s následnou modifikací radioaktivním izotopem na 5' konci. Zatímco dnes je pro specifické štěpení možné použít restriční endonukleázy, původně bylo nutné nejprve příslušné nukleotidy chemicky modifikovat a následně byla DNA v místě modifikace štěpena působením piperidinu. Modifikační reakce byly celkem čtyři- pyrimidiny (C+T) byly metylovány hydrazinem, přičemž po přidání NaCl byla tato reakce specifická pouze pro cytosin. Puriny byly depurinovány působením kyseliny mravenčí a guanin byl metylován dimetyl sulfátem. Rozštěpené fragmenty byly rozděleny na elektroforézovém gelu a k zobrazení byla využívána autoradiografie značených izotopů.

Přestože byla tato metoda v roce 1994 modifikována pro snadnější a použití (Boland et al. 1994), dnes se v podstatě nepoužívá.

1.3.2.2 Řetězcová metoda (Sanger sequencing)

Druhá metoda „první generace“ sekvenování byla vynalezena Frederikem Sangerem a až do nástupu sekvenačních metod nové generace byla nejdůležitější a nejrozvinutější sekvenační technikou na světě (Sanger and Coulson 1975, Sanger et al. 1977). Její princip je založen na DNA replikaci, při které jsou do nově vznikajícího řetězce DNA selektivně inkorporovány dideoxynukleotidy. Ty postrádají 3' hydroxylovou skupinu a proto je za nimi polymerační reakce zastavena. Sekvenace probíhá ve čtyřech samostatných reakcích vždy s jedním příslušným dideoxidem. Protože by čtení v případě úplného nahrazení jednoho ribonukleotidu bylo velmi krátké, je příslušný dideoxid do reakce pouze přidán a ribonukleotidy jsou v nadbytku. Polymerační reakce tak ve většině případů může dojít až do konce.

Původně byly dideoxidy značeny radioaktivně a analýza rozdělení fragmentů podle délky probíhala na elektroforézovém gelu. Postupně byla tato modifikace nahrazena fluorescenční značkou, která umožnila sekvenování všech čtyř ribonukleotidů v jedné reakci (Smith et al. 1986). Rovněž samotný postup sekvenace byl v průběhu let velmi zjednodušen- automatické sekvenátory využívající kapilární elektroforézu jsou v dnešní době schopné osekvenovat až 384 vzorků zároveň.

Jistou nevýhodou techniky je její neschopnost osekvenovat prvních zhruba 40 nukleotidů DNA za primerem. Výhodou jsou naopak poměrně dlouhé sekvenované úseky oproti jiným metodám (průměrně 800 bp, ovšem při vhodných podmínkách až 2000 bp) s malou chybovostí. Přestože byly vynalezeny novější a modernější techniky sekvenování, zůstává Sanger sekvenování nejlepší metodou pro sekvenaci repetitivních oblastí či klonárních inzertů bez nutnosti následné assembly.

1.3.2.3 Sekvenování „nové generace“ (Next Generation Sequencing, NGS)

NGS technologie jsou komerčně dostupné od roku 2004. Základní princip sekvenace je stejný jako v případě Sanger technologie- báze jsou identifikovány pomocí fluorescenčních značek zakomponovaných do DNA při její replikaci. Hlavní inovace metod spočívá v paralelním sekvenování velkého množství vzorků najednou, čemuž napovídá i jiný název- high throughput sequencing, tedy vysokorychlostní sekvenování. V posledních letech je vývoj těchto metod skutečně rapidní- počet čtení, které je možné získat z jedné paralelní reakce, každým rokem exponenciálně narůstá. Dnes je počet bází, které je možné získat z jednoho paralelního čtení více než 10^{12} (data Illuminy). Nové způsoby sekvenování také výrazně snížily jeho cenu a celogenomová sekvenace je v současnosti jen zlomkem ceny původní. V této práci uvádím pouze tři typy NGS, kromě první uvedené metody se jedná o nejznámější a nejpoužívanější metody vysokorychlostní sekvenace.

1.3.2.3.1 MPSS (Massively Parallel Signature Sequencing)

První metoda sekvenování nové generace byla vynalezena v devadesátých letech firmou Lynx Therapeutics. Základní princip metody spočíval v ligaci adaptérů a jejich následném čtení po čtyřech nukleotidech. Protože byla metoda velmi komplexní, nikdy nebyl vynalezen komerčně dostupný sekvenátor, nicméně charakterem čtení zcela určitě náleží k vysokorychlostním typům sekvenací (Brenner et al. 2000).

1.3.2.3.2 454- pyrosekvenování

Prvním komerčně dostupným sekvenátorem nové generace byl 454 Sequencer 20 od firmy Roche. Principem metody je paralelní pyrosekvenční reakce (Gharizadeh et al. 2006), probíhající ve stovkách tisících kapek obsahující homogenní fragmenty DNA. Každá kapka obsahuje průměrně jeden fragment DNA a k její amplifikaci dochází v olejovém roztoku v jamkách na destičce (Margulies et al. 2005). DNA se specifickými 454 adaptory je denaturována na jednovláknou, následně je amplifikována za inkorporace jednoho z nukleotidů. Při reakci dojde k odloučení pyrofosfátu, ten je začleněn do ADP za vzniku ATP a stává se funkčním přítomný enzym luciferáza (Froehlich et al. 2011). Ten produkuje viditelné světlo. Neinkorporované báze jsou degradovány apyrázou.

Délka čtení u 454 sekvenace vzrostla z původních 100-150 bp při 200.000 čtení na reakci (20M přečtených bází) na délku jednotlivých čtení až 800 bází a 14G přečtených bází (Huse et al. 2007). Nevýhodou metody zůstává poměrně vysoká cena.

1.3.2.3.3 Illumina (dříve Solexa) sekvenování

Druhá metoda NGS sekvenování byla poprvé představena roku 2006. K paralelizaci reakcí sekvenátoru Illuminy dochází při in situ amplifikaci fragmentů DNA, které jsou spojeny s pevně kotvenými adaptorovými sekvencemi. Dochází tak ke vzniku lokálních kolonií DNA, tzv. spojených DNA clusterů. Před sekvenováním je DNA knihovna denaturována na jednovláknou DNA. Ke zjištění sekvence jsou využívány čtyři typy terminačních bází ukončujících amplifikaci DNA. Fluorescenčně značené nukleotidy jsou snímány kamerou. Následně jsou značky odmyty a může probíhat další reakční cyklus. Na rozdíl od pyrosekvenování probíhá amplifikační reakce postupně- vždy dojde k připojení pouze jednoho nukleotidu (Bentley et al. 2008).

Původně bylo množství bází získaných z jednoho sekvenování touto technikou zhruba 1 Gb, ovšem postupným zdokonalováním techniky bylo dosaženo rychlosti čtení až 600Gb za 8 dní (jedná se o tzv. HiSeq2000 metodu). Chybovost je nižší než 2%. V porovnání s ostatními používanými typy je sekvenace Illuminou nejlevnější, předpokládá se, že by v budoucnosti mohla celogenomová sekvenace stát méně, než 1.000 \$.

1.3.3 Kompletace genomu (Assembly)

Poté, co je genom nějakého modelového organismu osekvenován, je nezbytně nutné ho zkompletovat. Sekvenační technologie neumožňují delší čtení než 15.000 bp (NGS sekvenování

Pacific Biosciences) (Eid et al. 2009), ve většině případů jsou ovšem jednotlivá čtení kratší- zhruba 2000 bází při použití Sanger technologie. U moderních typů sekvenování se délka ještě zkracuje- např. Illumina/Solexa má délku čtení maximálně 100-150 bází. Je jasné, že bez genové kompletace by orientace v genových sekvencích byla prakticky nemožná.

Pro sestavování genomů z jednotlivých čtení slouží bioinformatické programy vycházející z předpokladu, že se vysoce podobné DNA fragmenty nacházejí v genomu na stejné pozici. Sekvenční podobnost slouží jako místo pro spojení jednotlivých fragmentů k sobě a dojde k vytvoření větších úseků, kterým se říká contigy.

Problém mohou při sestavování genomu způsobit stejné či velmi podobné sekvence nacházející se v genomu na odlišných místech. Nejproblematictějšími bývají repetitivní sekvence, které jsou identické téměř v celém genomu a mohou být zdrojem stejných nebo velmi podobných fragmentů pocházejících z naprosto odlišných míst. Assemblační program může při sestavování takové místo chybně vyhodnotit a spojit i zcela nesousedící oblasti. To, jak moc bude program ovlivněn, závisí především na délce čtení- pokud je repetitivní oblast kratší, k chybnému spojení nedochází.

Teoreticky by tedy bylo lepší použít co nejdelší čtení, které poskytuje technologie od Pacific Biosciences. Tato technologie je čím dál tím víc využívána u bakteriálních genomů, u kterých je většina repetitivních sekvencí kratších než 6kbp. Značnou nevýhodou použití tohoto systému je nerovnoměrná délka čtení- zatímco nejdelší mají až 15kbp, většina jich je jen 800 bází dlouhých. Druhou nevýhodou je vysoká chybovost- až 15% bází je přečteno chybně (Wetzel et al. 2011). Důležitým prvkem pro správnou assembly je právě i předpokládaný počet chyb v jednotlivých čteních- čím je jich více, tím více musí program tolerovat možné sekvenční odchylky mezi spojovanými fragmenty a je tedy i vyšší pravděpodobnost spojení dvou nesousedních fragmentů.

Moderní sekvenační přístupy generují data se dvěma vlastnostmi: orientací a vzdáleností mezi jednotlivými bázemi (jedná se o tzv. mate-pair sekvenaci). Právě schopnost analyzovat a vyhodnotit tyto informace je nezbytná pro spojení contigů i přes repetitivní sekvence do velmi dlouhých scaffoldů. Přestože existuje velké množství assemblačních programů s touto schopností, ukazuje se, že pro sestavování do scaffoldů jsou nejlepšími programy ty, které se zabývají pouze problémem spojení contigů do velkých celků- tzv. scaffoldery. Existuje jich celá řada, v současné době jsou nejvyužívanější např. Bambus (Pop et al. 2004), SOPRA (Dayarian et al. 2010) nebo Opera (Gao et al. 2011). Tyto programy jsou úspěšně pravděpodobně kvůli využití dat z různých typů sekvencí.

Přestože existuje velké množství matematických algoritmů řešících problémy sestavování genomu, jsou i nadále nedokonalé a produkují velké množství chyb. Najít takovou chybu je ovšem velmi obtížné, především proto, že není známo, jak by měla správně sestavená assembly vypadat. Chyby se nenacházejí pouze ve správném seřazení jednotlivých čtení, ale i v sekvenci samotné. Zatímco velké chyby je možné najít např. srovnáním s BAC sekvencí nebo přímo fyzickým namapováním BAC klonu do genomu, najít jednonukleotidové záměny či lokální přestavby je prakticky nemožné.

1.3.3.1 Genom *Xenopus tropicalis*

1.3.3.1.1 Verze 4.1 (v.4.1)

Sekvenace genomu byla u druhu *X. tropicalis* zahájena Joint Genome Institutem v roce 2002 a pokračovala až do roku 2010. Získaná čtení byla postupně sestavována do několika verzí assembly. Jedna z nejvýznamnějších a dodnes používaných byla assembly verze 4.1, která byla zveřejněna v srpnu roku 2005. Přestože existuje i novější verze assembly, je stále používána pro přílišnou kontroverzi novější verze 7.1 (viz níže) a rovněž sloužila jako zdroj dat pro sestavení aktuální vazebné mapy (Wells et al. 2011).

K sestavení genomu *X. tropicalis* byla použita čtení, která jeho genom pokrývala přibližně 7,6x. Genom byl zkompletován programem JAZZ assembler (Aparicio et al. 2002) a obsahuje 19,501 scaffoldů s celkovou velikostí 1,51Gbp, do kterých je namapováno asi 97,6% všech známých cDNA. 272 největších scaffoldů pokrývá svou délkou přibližně polovinu genomu. Tato assembly je ovšem v porovnání s jinými genomy poměrně nekompletní - je sestavena zhruba z 175.000 souvislých contigů, které jsou mezi sebou propojeny neznámou sekvencí (představuje ji „N“). U některých scaffoldů počet N odpovídá až 25% celkového počtu bází.

1.3.3.1.2 Verze 7.1 (v.7.1)

Nejnovější verze genomu v7.1 byla představena v roce 2011 a dodnes nebyla oficiálně publikována. K jejímu sestavení byl použit program Arachne assembler (Batzoglou et al. 2002), dostupná data ze vznikající vazebné mapy (Wells et al. 2011) a rovněž syntenie s kurem domácím. Přibližně 70% genomu je obsaženo v deseti velkých superscaffoldech, které odpovídají chromozómům s příslušným číslem (Khokha et al. 2009). Rozdíly mezi posledními dvěma verzemi assembly ovšem nespočívají jen v sestavení menších scaffoldů z v.4.1. do

velkých superscaffoldů ve v.7.1. Data ze syntení vedla k rozdělení mnohých scaffoldů ze starší verze na části, které se v nové verzi genomu nachází v zcela jiných superscaffoldech, respektive chromozómech. Právě kvůli použití syntení, které nebyly vždy potvrzeny jinými přístupy, je tato verze genomu považována za velmi spornou.

1.3.3.2 Genom *Xenopus laevis*

V současné době probíhá sekvenování genomu i u druhu *Xenopus laevis* a jedná se o skutečně celosvětový projekt- spolupracují na něm univerzity z Texasu, Berkeley a Japonska. Data nebyla zatím publikována, ale současná verze genomu je dostupná na www.xenbase.org. Assembly zatím pokrývá přibližně 2Gbp, ovšem je velmi nedokonalá- celkový počet scaffoldů je přibližně 1 milion a navíc jsou vesměs velmi krátké (většinou kolem 10kb).

1.3.3.3 Zdroje informací

V posledních letech vzniklo velké množství internetových stránek, jejichž úkolem je shromažďovat dostupná data o obou modelových organizmech. Největší množství dat se nachází na www.xenbase.org, kde je možné zobrazit obě verze assembly *X. tropicalis* a nově i *X. laevis*, rovněž se zde nachází databáze cDNA, expresní databáze či informace o BAC knihovně. Veškerá data je možné stáhnout v aktuálních bioinformatických formátech.

Genové prohlížeče druhu *X. tropicalis* se nachází i na stránkách Joint Genome Institutu (<http://genome.jgi-psf.org/Xentr4/Xentr4.info.html>) či Ensemblu (http://www.ensembl.org/Xenopus_tropicalis/).

1.4 Separace chromozómů

Pro oddělení jednotlivých chromozómů od sebe je možné použít dvě zcela odlišné techniky- průtokovou cytometrii a mikrodisekci. Výhodou první ze zmíněných metod je možnost získání velkého množství chromozómů najednou. Značnou nevýhodou této techniky je ovšem její rozlišovací schopnost- pro správné oddělení je nezbytné, aby byl daný chromozóm výrazně větší či menší oproti ostatním. Doposud existuje pouze několik protokolů pro získávání jednotlivých chromozómů, vesměs u nejpoužívanějších modelových organismů.

Rozdělit chromozómy pomocí mikrodisekce je na druhou stranu možné vždy, když jsou rozlišitelné pod mikroskopem. Pro vyšší rozlišení je možné využít i různé pruhovací techniky. Rovněž je možné vyříznout pouze určitou část chromozómu či dokonce pruh. Zisk z jedné mikrodisekce je ovšem velmi malý, většinou se jedná o několik až několik desítek získaných

chromozómů. Z tohoto důvodu je pro následnou analýzu většinou nezbytná celogenomová amplifikace.

1.4.1 Mikrodisekce

Mikrodisekce je metoda sloužící k separaci velmi malých útvarů na úrovni buněk či buněčných kompartmentů z řezů tkání či buněčných kultur. Velmi často je využívána v případě, kdy je nutné oddělit jednotlivé typy buněk od sebe, např. pro proteinovou či genetickou analýzu. Existují dva rozdílné typy mikrodisekcí- starší, při které se využívaly velmi tenké jehly, a novější, využívající laserový paprsek (Berns and Floyd 1971).

1.4.1.2 Laserová mikrodisekce

Jak již bylo řečeno, při této technice je kombinována světelná mikroskopie s paprskovým laserem. Umožňuje izolaci konkrétních buněk a chromozómů z tkání či kultur. Buňky jsou naneseny na membránu umožňující snadné odmytí proteinů a zobrazeny pod mikroskopem. Ty, které mají být izolovány, jsou z tkání vyříznuty a přeneseny pomocí laserového paprsku.

Existují dva typy laserové izolace- využívající buď ultrafialové (UV) (Kölble 2000, Micke et al. 2005), nebo infračervené paprsky (IR) (Emmert-Buck et al. 1996). Výhodou systému využívajícího ultrafialové laserové paprsky je, že nedochází k zahřívání nebo zchlazování membrány s buňkami. Rovněž samotný přenos materiálu je na rozdíl od IR systémů nekontaktní- izolované buňky či chromozómy jsou laserovým paprskem katapultovány přímo do zkumavky (Ladanyi et al. 2006). V technologii využívající infračervené paprsky je před samotným transferem vytvořen komplex polymerů s izolovanými buňkami a proto je zvýšena pravděpodobnost kontaminace vzorku jiným materiálem. Ve forenzním výzkumu se proto používá častěji systém s ultrafialovými laserovými paprsky.

Značným problémem laserové mikrodisekce je malé množství získaného materiálu a rovněž jeho degradace při vyřezávání a přenosu. Největší množství DNA lze získat při izolaci chromozómů přímo do víčka sběrné mikrozkušavky. Při použití tohoto systému je rovněž snížena pravděpodobnost ztráty či degradace DNA, ke které dochází během promývání, eluci a přenášení vzorků mezi zkumavkami.

1.4.2 Průtoková cytometrie

Metoda je založená na principu oddělení různě nabitých částic v proudu kapaliny pomocí laserového paprsku. V dnešní době patří mezi nejpoužívanější metody průtokové cytometrie

fluorescenčně aktivovaná průtoková cytometrie (FACS), využívající fluorescenční značku na povrchu oddělované partikule. Pomocí průtokové cytometrie je možné oddělit 10^2 - 10^4 chromozómů za vteřinu (Doležel et al. 2012). Chromozómy jsou barveny chromozomálně-specifickým fluorochromem a jsou rozděleny podle intenzity fluorescence. Při prvních pokusech na chromozómech křeččíka čínského byl použit jako fluorochrom etidium bromid (Grey et al. 1979), který později nahradil Hoechst 33258 (Carrano et al. 1979). V současné době se využívá nejen intenzita, ale i množství fluorescenční barvy podél chromozómu (Bartholdi 1990, Rens et al. 1994).

Průtoková cytometrie chromozómů je využívána především v humánní medicíně. Existují protokoly pro jejich rozlišení jak při použití buněčných linií (van den Engh et al. 1984), ale i z periferní krve (Matsson and Rydberg 1981) a tkání (Kooi et al. 1984). I přes veškerý výzkum se ovšem doposud nepodařilo vyvinout protokoly pro všechny lidské chromozómy: chr. č. 9,10,11 a 12 zatím není možné získat odděleně. Průtoková cytometrie rostlinných chromozómů je shrnuta v (Doležel et al. 2012).

1.5 Celogenomové amplifikace (Whole Genome Amplification, WGA)

Množství získaného materiálu při laserové mikrodisekci je velmi malý, většinou je to pouze pár buněk či několik chromozómů. Velmi často je proto pro následnou analýzu nezbytná celogenomová DNA amplifikace. Tato technika je využívána i v preimplantační diagnostice, kdy jako zdroj DNA slouží třeba i jediná buňka.

1.5.1 PEP (Primer Extension Preamplification)

PEP je nejstarší celogenomovou amplifikací (Zhang et al. 1992). Jde vlastně o PCR reakci, ve které nasedají náhodné primery o délce 15 bází na templátovou DNA při teplotě 37°C. V takto nízké teplotě mohou primery nasedat na templát nespecificky a tím se zvyšuje množství naamplifikované DNA. Ta je replikována Taq polymerázou při teplotě 55°C. Modifikací této metody je I-PEP (Improved- PEP), která využívá krom Taq polymerázy také Pwo polymerázu a do každého cyklu je přidán elongační krok při teplotě 68°C (Dietmaier et al. 1999). Oproti původnímu typu je I-PEP v celogenomové amplifikaci účinnější, ovšem v dnešní době byly obě tyto metody nahrazeny modernějšími postupy.

1.5.2 DOP-PCR (Degenerate Oligonucleotide Primed- Polymerase Chain Reaction)

Metoda DOP-PCR je založena na nesedání částečně degenerovaných primerů, které se během nízkoteplotních cyklů váží na mnoha místech v genomu. Specifického anealingu primerů je možné dosáhnout zvýšením teploty. Množství templátové DNA může být při DOP-PCR v rozmezí 15pg-400 ng, přičemž množství získané DNA je vyšší než při použití techniky PEP (Wells et al. 1999, Peng et al. 2007). Hlavní výhodou techniky je možnost produkce dostatečného množství DNA pro genovou hybridizaci i z jedné buňky (Wells et al. 2002). Pouze částečně zdegenerované primery ovšem příčinou nenáhodné amplifikace DNA, což značně ovlivňuje některé molekulárně biologické testy.

Délka amplifikované DNA je přibližně stejná jako při PEP (Telenius et al. 1992). Modifikace DOP-PCR umožňuje i produkci delších amplifikátů (přibližně 0,5- 10 kb dle použitého protokolu) (Kittler et al. 2002). Jedná se o LP-DOP-PCR (Long Products DOP-PCR), která je založena na exonukleázové funkci Pwo polymerázy a delších časů v cyklech během PCR amplifikace.

Další možnou modifikací DOP-PCR je využití více degenerovaných primerů a 12 nespecifických amplifikačních cyklů (Bonnette et al. 2009). Pro zvýšení délky amplifikátů jsou do reakce přidávány další typy polymeráz. Tato modifikace vede oproti původní metodě k produkci delších fragmentů DNA s pokrytím genomu vyšším až o 45%.

1.5.3 MDA (Multiple Displacement Amplification)

Třetím přístupem pro celogenomovou amplifikaci je MDA. Jejím principem je anealing krátkých, pouze šestinukleotidových primerů na jednovláknou DNA. Při jejich prodlužování DNA polymerázou dochází k odštěpování nově vzniklých fragmentů, které mohou rovněž sloužit jako templáty, ovšem v opačné orientaci. Při MDA tedy vzniká velké množství kopií zdrojové DNA (Lovmar and Syvänen 2006). Jako polymeráza slouží nejčastěji phi fágová polymeráza $\phi 29$, případně Bst DNA polymeráza, která ovšem produkuje větší množství chyb (Spits et al. 2006).

V klinické praxi je MDA amplifikace používána např. v prenatalní diagnostice pro zjištění některých genetických vad, jako jsou X-vázaný retinoblastom (Lledó et al. 2008) či syndrom fragilního chromozómu X (Malcov et al. 2007). MDA je rovněž možné použít pro získání velkého množství DNA fragmentů s vysokým pokrytím genomu z jedné spermie (Jiang et al. 2005).

1.5.4 OmniPlex WGA

Poslední metoda celogenomové amplifikace využívá teplotní fragmentaci DNA na definovanou velikost. Na tyto náhodné fragmenty jsou připojeny sekvence sloužící pro nasednutí primerů pro následující amplifikaci. Těmto fragmentům s primery se říká knihovna a může být mnohonásobně (třeba i milionkrát) namnožena. Hlavní výhodou této metody je možnost současného testování mnoha lokusů v nezávislých PCR reakcích.

OmniPlex WGA je v současné době jednou z nejpoužívanějších typů celogenomové amplifikace a rovněž jsou na jejím principu založeny některé z komerčně dostupných kitů.

1.5.5 Nevýhody celogenomových amplifikací a jejich srovnání

Při žádném z typů celogenomových amplifikací se doposud nepodařilo naamplifikovat kompletně celý genom. Důvody k tomu jsou zřejmé- templátové DNA je velmi malé množství, a tamůže být po některých typech izolace i poškozená. Další příčinou je metodika samotná. I když jsou primery velmi krátké a často degenerované, některé fragmenty jsou oproti jiným množeny častěji. Často také dochází ke ztrátě alel či mikrosatelitů používaných v různých studiích.

Studie (Höckner et al. 2009) ukázala, že z dostupných metod má největší množství amplifikovaných mikrosatelitů komerčně dostupný kit GenomePlex Single Cell Kit firmy Sigma-Aldrich, využívající metodu OmniPlex WGA. V této studii sloužily jako zdroj DNA mikrodisektované chromozómy a byly porovnávány všechny čtyři typy celogenomové amplifikace.

1.6 Bioinformatika

Bioinformatika je v současnosti nedílnou částí biologických analýz. Především neustálý nárůst počtu sekvencí získaných ze sekvenací nové generace vede k nezbytnosti využití bioinformatických metod pro třídění a vyhodnocení dat. Existuje mnoho volně dostupných i komerčně prodávaných programů, ty ovšem neumí řešit všechny problémy současně. Pro bioinformatika je z tohoto důvodu velmi často nutná alespoň elementární znalost nějakého programovacího jazyka, který mu umožní upravit si program přesně pro své účely.

1.6.1 Algoritmy pro vyhodnocování dat ze sekvenací

Pro vyhodnocení dat ze sekvenací existuje nepřehledné množství algoritmů a rovněž je možné naprogramovat si nějaký svůj vlastní. Vzhledem k množství dat získaných především z NGS sekvenací je ovšem naprosto nezbytné, aby takové programy pracovaly co nejúsporněji, tj.

aby byly co nejlépe přizpůsobené konkrétnímu problému. Velmi často jsou výstupem z NGS sekvenací desítky až stovky miliónů krátkých sekvencí, které je nutné buď spojit (assemblovat), anebo srovnat s assembly již existující. Při použití nevhodného softwaru by taková operace mohla trvat týdny až měsíce, a to i při použití velmi výkonného počítače.

Mezi nejpoužívanější algoritmy pro aligning čtení do genomu patří SOAP2 (Li et al. 2008b), Bowtie (Langmead et al. 2009), BWA (Li et al. 2008a), MAQ (<http://maq.sourceforge.net>) a Novoalign (www.novocraft.com). Kromě algoritmu MAQ je u všech nutné nejprve vytvořit indexovaný referenční genom, se kterým jsou následně jednotlivá čtení srovnávána. Každý z těchto programů používá jiný systém pro vyhledávání zásahů, Novoalign má od ostatních i odlišný způsob vytváření indexu. Výhodou algoritmů SOAP2, Bowtie a Novoalign je rovněž možnost zvolit maximální počet nepřesných zásahů oproti referenčnímu genomu. Testování těchto programů nicméně ukázalo, že při použití kvalitních sekvencí mezi nimi v podstatě neexistuje rozdíl a všechny tyto programy fungují stejně dobře (Yu et al. 2012). Nejpoužívanějším algoritmem je v současné době pravděpodobně Bowtie, který je neustále vyvíjen a především je spojen s dalším nepřeborným množstvím programů pro srovnávání sekvencí nejen s DNA, ale i m- RNA a genů, a dále i s programy na vytváření scaffoldů, assembly, a rovněž mnoha dalšími (<http://bowtie-bio.sourceforge.net/>).

1.6.2 Datové formáty pro zaznamenání sekvence

Různé programy požívají různé formáty dat. Většinou se jedná o zjednodušené zapsání sekvence, orientace a jejich dalších vlastností. Jedna přečtená sekvence se často nachází na jednom řádku, případně je od ostatních sekvencí oddělena jiným způsobem (např. specifickým znakem). Mezi nejznámější formáty patří FASTA, případně FASTQ. Každý program používá trochu jiný datový formát a velmi často je nezbytné tyto formáty mezi sebou převádět. Existuje jich poměrně velké množství a i některé internetové databáze mají svoje vlastní formáty. Univerzální formát nebyl zatím vytvořen. Z tohoto důvodu se budu zabývat pouze nejznámějšími formáty a uvedu je spíše pro ilustraci.

1.6.2.1 FASTA

Nejjednodušší a nejpoužívanější formát, ve kterém každá sekvence začíná „>“, následuje název, případně popis a na dalším řádku je uvedena samotná sekvence. Tento formát je využíván v internetových vyhledávačích, případně jako výstup ze Sanger sequencing.

1.6.2.2 FATSQ

Tímto formátem je kromě sekvence zaznamenávána rovněž kvalita sekvence při výstupu ze sekvenace. Sekvence začíná „@“, názvem, na dalším řádku je samotná sekvence a na třetím řádku je zaznamenána kvalita sekvence, kterou je možné vyhodnotit v některých programech.

1.6.2.3 Další formáty databází

Velké množství datových formátů využívá systém jedna sekvence-jeden řádek. Každý řádek obsahuje krom sekvence různé množství informací, které jsou od sebe odděleny mezerou či tabulátorem. V každém formátu je počet jednotlivých informací pevně daný, a pokud není tato informace známá, je místo ní zobrazena tečka. Jednotlivé formáty si jsou vesměs velmi podobné- pro zjednodušení bude podrobně popsán pouze jeden. Krom formátu BED existují i formáty BAM, SAM, MySQL, GFF a mnoho dalších. V těchto formátech jsou uloženy informace o známých, již srovnaných sekvencích a může se jednat jak o sekvence DNA, RNA i proteinové sekvence.

1.6.2.3.1 BED formát

Tento formát obsahuje tři povinné položky, zbylých devět je dobrovolných. Je používán některými zobrazovacími softwary, např. programem GenomeBrowse. Jednotlivé položky jsou od sebe odděleny tabulátorem a na jednom řádku se vždy nachází informace o jedné sekvenci.

Jedná se o následující položky:

1. Scaffold či chromozóm, na kterém se sekvence nachází
2. První pozice sekvence (Start)
3. Poslední pozice sekvence (End)
4. Jméno
5. Skóre
6. Vlákno- „+“ nebo „-“
7. Druhý start- např. start kodón
8. Druhý konec- např. stop kodón
9. ItemRGB- barva v zobrazovacím programu
10. Počet bloků- počet exonů v jedné BED linii
11. Velikost bloků, které jsou mezi sebou odděleny středníkem
12. Začátky jednotlivých bloků

1.6.3 Základní programovací jazyky používané v bioinformatice

Pro bioinformatiku je vhodný jakýkoli programovací jazyk. Vzhledem k tomu, že biologové většinou nejsou programátoři, je víc než vhodné, aby byl jazyk co nejjednodušší a nejpřehlednější. V současné době patří mezi nejpoužívanější skriptovací jazyky PERL, PYTHON a případně trochu komplikovanější Java. Dalším používaným jazykem, především pro řešení statistických problémů, je R. Rovněž je vhodná znalost některého Unixového (Linux, BDS, Solaris) operačního systému.

1.6.3.1 PERL

Jedná se o pravděpodobně nejpoužívanější jazyk v bioinformatice. Byl vyvinut v roce 1987 Larrym Wallem a jeho hlavní výhodou je jednoduchost, flexibilita a rovněž snadná instalace modulů. Je volně dostupný na www.perl.org. Pro jeho interpretaci je nezbytné mít nainstalovaný další program, tzv. interpreter. Pro bioinformační účely bylo vytvořeno velké množství modulů, které je možné stáhnout si z internetu. Souhrnně se tato modifikace PERLU jmenuje BioPERL (www.bioperl.org). I mnoho dalších bioinformačních programů používá PERL, například GBrowse (<http://gmod.org/wiki/GBrowse>), který je běžně používaný na mnoha stránkách pro zobrazení assembly, genů a sekvence.

1.6.3.2 PYTHON

Druhý nejpoužívanější jazyk, který má na rozdíl od PERLU pevně danou skladbu. Z tohoto důvodu je doporučován pro začátečníky. Na druhou stranu u něj neexistuje tak velké množství modulů jako u PERLU a rovněž jejich instalace je složitější, protože neexistuje centrální úložiště, ze kterého by bylo možné tyto moduly stáhnout. Stejně jako PERL je i PYTHON volně dostupný (www.python.org).

2. Cíle práce

Hlavním cílem práce bylo nalézt nové metody pro genetické mapování u modelového organismu *Xenopus tropicalis*.

Cíle jednotlivých studií byly následující:

- Sestavení genetické mapy druhu *Xenopus tropicalis* se zaměřením na fyzické mapování genů ve vybraných scaffoldech pro přiřazení jednotlivých LG k chromozómům a zjištění jejich orientace
- Optimalizace laserové mikrodisekce chromozómů *X. tropicalis*, celogenomové amplifikace a jejich použití pro Zoo-FISH značení chromozómů *Xenopus laevis*
- Identifikace scaffoldů a genů na krátkém raménku chromozómu 7 novým postupem využívajícím mikrodisekci příslušné chromozomální oblasti a její sekvenování metodou Illumina
- Charakterizace rozdílů pohlavních chromozómů u *X. tropicalis* s využitím dat získaných sekvenací mikrodisektovaného p raménka chromozómu 7

3. Výsledky

Developmental Biology 354 (2011) 1–8



Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology



Resource

A genetic map of *Xenopus tropicalis*

Dan E. Wells^a, Laura Gutierrez^a, Zhenkang Xu^a, Vladimir Krylov^b, Jaroslav Macha^b, Kerstin P. Blankenburg^c, Matthew Hitchens^c, Larry J. Bellot^a, Mary Spivey^a, Derek L. Stemple^d, Andria Kowis^c, Yuan Ye^c, Shiran Pasternak^c, Jenetta Owen^a, Thu Tran^a, Renata Slavikova^b, Lucie Tumova^b, Tereza Tlapakova^b, Eva Seifertova^b, Steven E. Scherer^c, Amy K. Sater^{a,*}

^a Department of Biology and Biochemistry, University of Houston, Houston TX USA

^b Department of Cell Biology, Charles University in Prague, Faculty of Science, Czech Republic

^c Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX USA

^d Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1HH, UK

ARTICLE INFO

Article history:

Received for publication 24 April 2010

Revised 5 March 2011

Accepted 11 March 2011

Available online 31 March 2011

Keywords:

Xenopus

X. tropicalis

Genetic map

Genome

Simple sequence length polymorphism

ABSTRACT

We present a genetic map for *Xenopus tropicalis*, consisting of 2886 Simple Sequence Length Polymorphism (SSLP) markers. Using a bioinformatics-based strategy, we identified unique SSLPs within the *X. tropicalis* genome. Scaffolds from *X. tropicalis* genome assembly 2.0 (JGI) were scanned for Simple Sequence Repeats (SSRs); unique SSRs were then tested for amplification and polymorphisms using DNA from inbred Nigerian and Ivory Coast individuals. Thus identified, the SSLPs were genotyped against a mapping cross panel of DNA samples from 190 F2 individuals. Nearly 4000 SSLPs were genotyped, yielding a 2886-marker genetic map consisting of 10 major linkage groups between 73 and 132 cM in length, and 4 smaller linkage groups between 7 and 40 cM. The total effective size of the map is 1658 cM, and the average intermarker distance for each linkage group ranged from 0.27 to 0.75 cM. Fluorescence *In Situ* Hybridization (FISH) was carried out using probes for genes located on mapped scaffolds to assign linkage groups to chromosomes. Comparisons of this map with the *X. tropicalis* genome Assembly 4.1 (JGI) indicate that the map provides representation of a minimum of 66% of the *X. tropicalis* genome, incorporating 758 of the approximately 1300 scaffolds over 100,000 bp. The genetic map and SSLP marker database constitute an essential resource for genetic and genomic analyses in *X. tropicalis*.

© 2011 Elsevier Inc. All rights reserved.

Introduction

X. tropicalis has emerged as a promising model system for genetic analyses of vertebrate development, extending the molecular, biochemical, and embryological strengths known from *Xenopus laevis* with the addition of genetic tractability. Initial genetic screens have yielded unique and interesting phenotypes (Abu-Daya et al., 2009; Goda et al., 2006; Grammer et al., 2005; Noramly et al., 2005), and the *X. tropicalis* genome assembly is now available (Hellsten et al., 2010). The integration of genetic and genomic approaches requires a genetic map for use in positional cloning of genes identified by mutation.

Although several types of polymorphic markers can be used to establish a genetic map, we chose Simple Sequence Length Polymorphisms (SSLPs), identified from an initial set of Simple Sequence Repeats (SSRs). Short, tandemly repeating sequences, SSRs are present at an average frequency of 1/20–40 kb (Strachan and Read, 2003); strain-specific differences in the length of the repeat are

reflected as differences in the size of a PCR fragment that encompasses the SSR.

SSLPs offer several advantages. Unlike Amplified Fragment Length Polymorphisms (AFLPs), SSLPs can provide a fully informative PCR assay. Although Single Nucleotide Polymorphisms (SNPs) have the potential for a significantly greater density, they are generally obtained by comparing two genomic sequences from different strains of the same species, thus requiring significantly more sequencing and assembly to permit identification. Second, candidate SSLPs can be identified from genomic sequence using a bioinformatics-based strategy, dramatically improving the ease and cost-effectiveness; moreover, as sequence assembly progresses, the search for SSLPs can be targeted to sparsely covered regions to improve uniformity of coverage. SSLPs have been used as the basis for genetic linkage maps in many vertebrate species.

Although historically SSLPs have been identified via cloning from small-insert genomic DNA libraries, we have used bioinformatics to identify SSRs within the *X. tropicalis* genome. These SSRs were then tested for polymorphisms between the Nigerian (N) and Ivory Coast (IC) strains of *X. tropicalis*. The resulting SSLPs were then genotyped on a Nc × IC F2 mapping panel to generate a 2886-marker genetic map.

* Corresponding author at: Dept. of Biology and Biochemistry, University of Houston, 4800 Calhoun Rd, Houston TX 77204-5001, USA. Fax: +1 713 743 2636.

E-mail address: asater@uh.edu (A.K. Sater).

Preliminary versions of this map and a database of the SSLP markers have been made available through our website (<http://tropmap.biology.uh.edu>).

Materials and methods

Map cross and F2 DNA panel

The map cross was derived from a single mating of two F1 progeny produced by a P1 cross of a male Nigerian (N) inbred over 7 generations (N7) × a female Ivory Coast (strain IC; inbred from original Adiopoudome collection) inbred over 5 generations (IC5); this initial P1 cross was carried out in the laboratory of Robert Grainger at the University of Virginia. The IC animals represent the same strain as that used in Khokha et al. (2009). The F2 progeny were euthanized as postmetamorphic tadpoles in preparation for DNA isolation. The DNA isolation protocol was modified from Sparrow et al. (2000). Briefly, metamorphic tadpoles were euthanized in 0.05% benzocaine, frozen individually in an ethanol/dry ice bath, and stored at -80°C . Each individual was macerated extensively, and the tissues were digested in 100 $\mu\text{g}/\text{ml}$ Proteinase K at 55°C overnight. Following heat-inactivation of Proteinase K, samples were treated with 50 $\mu\text{g}/\text{ml}$ RNase A at 37°C for 90 min. DNA was isolated after phenol–chloroform extraction and ethanol precipitation. The DNA isolation protocol can be found at <http://tropmap.biology.uh.edu/DNAisolation.html>. At least 300 μg of DNA was isolated from each individual. A DNA panel representing 190 F2 individuals as well as the 2 P1 individuals was used to map the SSLPs.

Bioinformatic identification of unique SSRs

Simple Sequence Repeats (SSRs) were identified within scaffolds from the JGI *X. tropicalis* genome assembly 2.0. Assembly 2.0 was the current assembly at the time this project was initiated, and it consists of over 27,000 scaffolds. SSRs were identified bioinformatically using an algorithm modified from Tandem Repeats Finder (Benson, 1999), which selected a single di-, tri-, or tetranucleotide SSR with unique flanking sequences (“unique SSR”) from each scaffold.

Assessment of polymorphisms

Primers were designed via Primer 3 (Rozen and Skaletsky, 2000; <http://www-genome.wi.mit.edu/cgi-bin/primer/primer3-www.cgi>) to amplify the SSR and flanking sequence to generate an amplified fragment between 80 and 500 bases under standard conditions. The amplified fragment was then tested via BLAST (McGinnis and Madden, 2004; <http://www.ncbi.nlm.nih.gov/BLAST>) to determine whether it represented a unique sequence within the genome. All primers were obtained from Illumina (San Diego, CA). Primer sets were tested initially for robust amplification of a single fragment visualized by agarose gel electrophoresis, using the following standard conditions (30 cycles, 58°C annealing temperature, and 1.5 mM MgCl_2). SSR primer sets showing robust amplification were then tested for polymorphisms using DNAs from unrelated Nigerian and Ivory Coast individuals as well as a pool of F2 DNAs represented in the map cross. Polymorphism PCR assay reactions included ^{32}P -dCTP, and radiolabeled PCR products were visualized following polyacrylamide gel electrophoresis. Sequence length polymorphisms were detectable as differences in the size of the PCR product; these assays also provided an estimate of allelic size differences.

Genotyping and linkage analysis

Genotyping of identified SSLPs against the F2 mapping panel was carried out via PCR reactions run in duplex, using primers labeled with the fluorochromes FAM and HEX (Illumina, San Diego, CA). PCR

products were separated on an ABI 3730 DNA Sequencer. Initial genotyping data were examined using GeneMapper 3.7 software (Life Technologies, Carlsbad CA). Linkage analysis was carried out using JoinMap 3.0 (Van Ooijen and Voorrips, 2001), which can incorporate a wide range of polymorphism types; thus, genotyping results were classified according to categories of polymorphism, as specified by JoinMap 3.0 protocols. Genotyping results were evaluated individually; markers that could not be definitively genotyped were eliminated from the analysis. For comparative purposes, a second linkage analysis was performed using MSTmap (Wu et al., 2008).

Preparation of metaphase spreads and Fluorescence In situ Hybridization (FISH)

Metaphase spreads were prepared from primary cell cultures derived from testes of juvenile frogs (Ivory Coast strain). Mitotic arrest was induced by treatment with 0.2 $\mu\text{g}/\text{ml}$ colchicine (Sigma) for 4.5 h at 28°C in 7% CO_2 atmosphere. Cells were collected by trypsinization and centrifugation at $100\times g$ for 10 min at room temperature (RT). Hypotonic shock, fixation and spreading of cells were described previously (Krylov et al., 2007). Slides were dried and treated with 50 $\mu\text{g}/\text{ml}$ pepsin in 0.01 N HCl for 5 min at 37°C , followed by a 30-minute incubation in 2% paraformaldehyde at RT. Endogenous peroxidase was quenched by a 30-minute incubation in 1% hydrogen peroxide. Methods for FISH-TSA, including preparation and labeling of cDNA probes, have been described previously (Krylov et al., 2007).

Assignment of linkage groups to *X. tropicalis* chromosomes

Assignment of linkage groups (LG) to *X. tropicalis* chromosomes was based on physical mapping of cDNA probes by means of fluorescence *in situ* hybridization coupled with a tyramide amplification step (FISH-TSA) (Khokha et al., 2009; Krylov et al., 2007). Appropriate genes or cDNA clones that were a minimum of 900 bp in length were selected from scaffolds anchored in respective linkage groups using Ensembl (<http://www.ensembl.org/index.html>) or GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>) databases. On one occasion (LG6, scaffold 104), the Gurdon Institute Xt. EST database (<http://genomics.nimr.mrc.ac.uk/cgi-bin/public.exe>) was used as a data source (see Supplementary Table 2). Chromosome identification and numbering were based on the p/q ratio following the revised nomenclature system described in Khokha et al. (2009). The position of the FISH-TSA signal was expressed as a relative distance from the centromere (RDC) with respect to the appropriate arm. Both the p/q arm ratio and RDC were determined using ACC Program v. 5.0 (SOFO, Brno, Czech Republic). At least six mitoses showing a FISH-TSA signal were evaluated for each physically mapped locus.

Results

Identification of SSLPs

Over 14,000 SSRs were identified using our data mining script. Several rounds of script modification were required to ensure that the selected markers represented unique sequences. Initially, the script selected only tri- and tetranucleotide repeats; once those were exhausted, the script was modified to include dinucleotide repeats, which are considerably more abundant (Xu et al., 2008).

Over 12,000 SSR primer sets were tested for amplification; of these, 9369 (78%) were robustly amplifiable under our standard conditions. The first 6400 primer sets included primarily tri- and tetranucleotide SSRs. At least one unique tri- or tetranucleotide SSR was identified from all scaffolds containing these sites. Scaffolds containing one or more unique tri- or tetranucleotide SSRs comprise approximately 85% of the genome. The remaining scaffolds were thus

not represented within this initial set. The remaining SSRs contain dinucleotide repeats; these are considerably more abundant and amplify more reliably than the tri- and tetranucleotide SSRs.

Unique SSR sequences were amplified from genomic DNA isolated from the Nigerian and Ivory Coast grandparents (P1 individuals). In some instances, DNA from unrelated Nigerian and Ivory Coast individuals were also included, or DNA from pooled F2 individuals was used instead. Out of 9300 SSRs tested, 5829 (63%) showed polymorphisms between the Nigerian and Ivory Coast individuals, and 4779 SSRs (82% of polymorphic SSRs) were polymorphic between the P1 individuals. A marker database, which includes primer sequences, sequence of the amplified fragment, SSR type, and scaffold assignment, is available at <http://www.tropmap.uh.edu>. A diagram of the workflow for identification of markers and generation of the genetic map is shown in Fig. 1.

The F2 map cross panel

A mapping cross F2 DNA panel representing 190 sibling individuals was generated from a single cross of two F1 progeny produced by the P1 cross of a male Nigerian (N) F7 with a female Ivory Coast (IC) F5. A total of 3982 SSRs (83% of the P1 polymorphic set) were genotyped using this DNA panel and DNA samples from each parent; these SSRs had been found to be polymorphic in either the P1 individuals or a pool of the F2 individuals. We were unable to place 1003 (25%) of the genotyped markers on the map; most of the unplaced markers were either difficult to score, did not show polymorphisms in the mapping cross, or were eliminated in preliminary iterations of the map because of anomalies in genotyping or linkage analysis. Of the 2979 polymorphic markers that could be clearly scored, only 93 (3%) were not resolvable on the map; presumably, these markers were too far away (>20 cM) from other

clearly scorable markers to show linkage. The mapping cross represents 380 informative meioses, and thus the resulting map has a theoretical limit of resolution of 0.26 cM.

Genotyping analysis

Primary linkage analysis was carried out using JoinMap 3.0 (Van Ooijen and Voorrips, 2001). JoinMap uses maximum likelihood to calculate recombination frequencies between all possible pairs of markers; from this data set, it assigns linkage groups, predicts marker order within linkage groups, and calculates the intermarker distance (Stam, 1993). JoinMap is particularly suitable for our studies because it permits a greater range of heterozygosity in parental genotypes, which allows us to use as many of the SSLPs as possible. It has been used extensively to create genetic linkage maps in plants (e.g., Kuittinen et al., 2004; Song et al., 2004). After an initial analysis that divided the markers into linkage groups, each linkage group was analyzed individually; upon completion of a linkage analysis in Joinmap, the results were compared with earlier versions of the map (i.e., preliminary versions prepared with fewer markers), and the genotyping results for selected markers were reviewed individually. In most cases, several markers were eliminated from a given linkage group, and the remaining markers were re-analyzed. With the exception of LG1, each of the major linkage groups was analyzed more than once.

Linkage analysis yielded a genetic map of 2886 SLP markers, organized in 10 major and 4 minor linkage groups. The linkage groups included 103–506 markers, in groups ranging from 77 to 138 centimorgans (cM). The LOD scores at which these linkage groups appeared ranged from 3 to 12. The 4 minor linkage groups included a total of 54 markers in groups ranging from 6 markers over 7 cM to 29 markers across 40 cM. A comparison of the linkage groups in terms of size, number of markers, and minimal genomic representation is shown in Table 1. A spreadsheet of the complete genetic map is

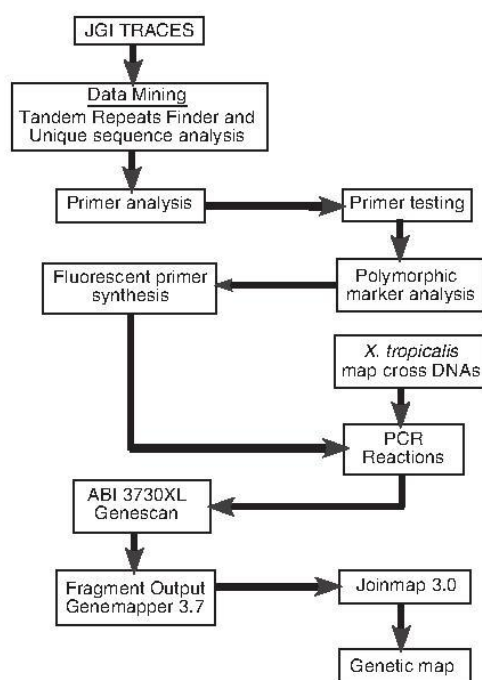


Fig. 1. Organization of Bioinformatics, SLP testing, and genotyping. Data mining of sequence traces from JGI genome assembly 2.0 identified SSRs, which were then tested for PCR amplification and polymorphisms. SSRs shown to be polymorphic in the initial P1 cross were used to genotype the map cross panel using an ABI 3730 XL sequencer. Genotyping results were scored using Genemapper 3.7, and linkage analysis was carried out using JoinMap 3.0.

Table 1
Linkage Groups in 2886-Marker Map.

LG	Chr. ^a	Old LG ^b	Total cM ^c	# markers	# scaffolds ^d	Genomic representation ^e (kB)	Avg intermarker dist (cM)
1	1	1	138.55	506	141	164,091.4	0.27
2	2	6	109.69	287	79	96,509.8	0.38
3	3	8	79.02	264	70	93,694.2	0.30
4	4	7	123.95	365	90	121,131.2	0.34
5	5	9	88.86	207	56	67,565.4	0.43
6	6	2	94.97	357	87	120,335.4	0.27
7	7	4	120.11	303	99	92,423.4	0.40
8	8	5	73.06	264	77	91,586.8	0.28
9	9	3	112.9	178	50	68,777.6	0.64
10	10	10	77.56	99	30	28,636.7	0.75
5b	5	A	39.96	29	15	81,869.5	
8c	8	E	16.36	12	8	3,978.0	
3b	3	F	7.49	8	5	5,881.6	
8b	8	5b	15.16	7	4	4,407.7	
Add'l ^f	–	–	–	–	45	87,379.0	
Total			1097.64	2886	856	1,128,267.7	

^a Chromosomes are numbered as described in Khokha et al. (2009).

^b "Old LG" refers to the initial linkage group designation previously shown on our website. All linkage groups are now numbered in accordance with the corresponding chromosome.

^c "Total cM" refers to the distance from the first to the last marker within a linkage group.

^d Only those scaffolds that are represented entirely within a single linkage group are included in the "# of scaffolds represented" for each linkage group.

^e "Genomic representation" refers to the combined size of all scaffolds represented within a linkage group; this number is a minimum estimate of the portion of the genome represented within a linkage group.

^f The total minimal genomic representation includes approximately 87 MB from scaffolds that appear on 2 or more linkage groups ("Add'l").

presented in Supplementary Table 1; this spreadsheet also includes the primers and amplified sequence for each marker, as well as the corresponding scaffold from *X. tropicalis* genome assembly 4.1 (Hellsten et al., 2010).

The total genetic interval across all linkage groups is 1097.64 cM. Since a mutation can be mapped if it is within 20 cM of a marker, the effective size of the map includes not just the total genetic interval, but also an additional 20 cM at either end of each linkage group (Knapik et al., 1996). Thus, the effective size of the map is approximately 1658 cM.

Assignment of linkage groups to chromosomes

Fluorescence *in situ* hybridization coupled with a tyramide amplification step (FISH-TSA) was used for the physical mapping of cDNA probes in order to assign linkage groups to *X. tropicalis* chromosomes. In total, 69 cDNA markers were localized (Fig. 2). A detailed description of mapped loci is provided in Supplementary Table 2. Several DNA probes were used for each LG, and probes for a single LG mapped to only one chromosome. A gynogynetic mapping strategy has previously been used to identify the locations of the centromeres and markers corresponding to centromeric positions (Khokha et al., 2009). LGs 2, 3, 4, 7, and 8 showed a p–q orientation, while the remaining linkage groups were oriented from q to p. All of the minor linkage groups were assigned to chromosomes; they are

listed as LG3b, 5b, 8b, and 8c. Assignments of 3b, 5b, and 8b were based on the position of probes representing scaffolds included solely within those minor linkage groups; assignment of 8c is based on the position of a marker from Scaffold 427, which is represented both in LG8 and in 8c.

With the exception of LG2, all linkage groups covered the length of the corresponding chromosomes (Fig. 3). Cytogenetic mapping of LG2 showed that this linkage group covers the q arm of Chr. 2, terminating near the centromere. The p arm of Chr. 2 is thus not represented in this genetic map.

Several scaffolds have been cytogenetically mapped to the p arm of Chr. 2; these will be described elsewhere (Mácha et al., manuscript in preparation). Markers representing these scaffolds were identified from our marker dataset, and the genotyping results for these markers were assessed. Of the genotyped markers from these scaffolds, 73% were nonpolymorphic in our F2 mapping panel. To establish a size-matched “control” dataset, we also identified markers from the scaffolds that flanked each of the Chr.2 p arm scaffolds in size, i.e., if scaffold 40 was represented on the Chr. 2 p arm, scaffolds 39 and 41 were included in the control dataset. We then reviewed the genotyping results for all markers from scaffolds in the control dataset; only 11% of these markers were nonpolymorphic across the F2 mapping panel, which is significantly different ($p < 0.0001$, Fisher's Exact Test) from that found for the markers representing the Chr.2 p arm scaffolds. These findings suggest that the level of polymorphism

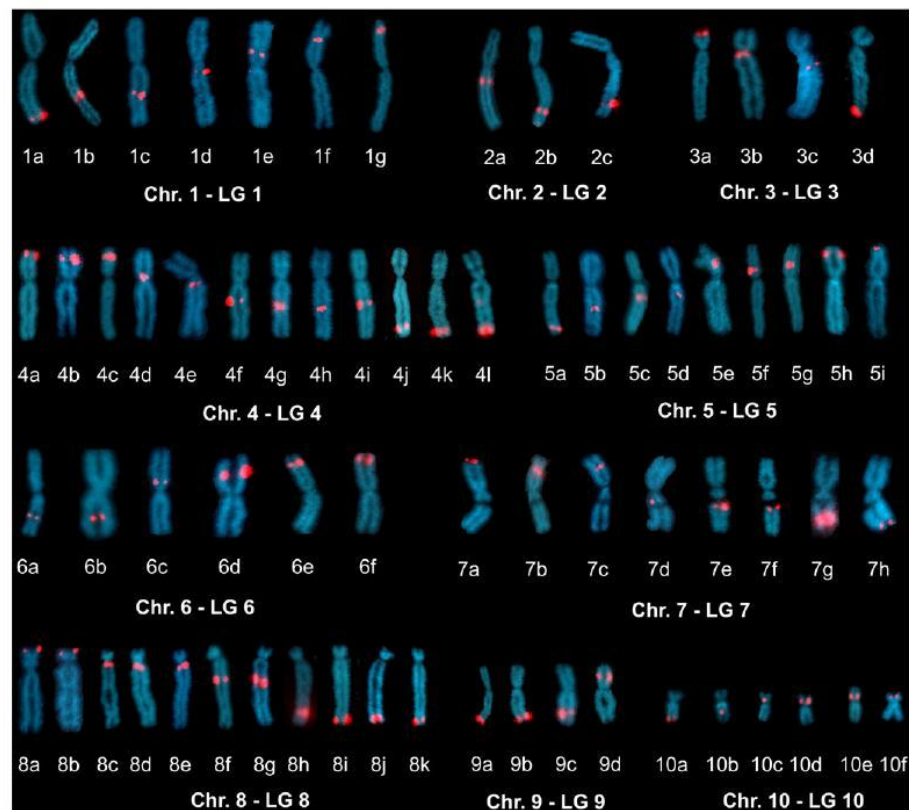


Fig. 2. Assignment of linkage groups to *Xenopus tropicalis* chromosomes via FISH-TSA using following cDNA probes. **LG1:** 1a. *pias2* (q 0.95), 1b. *cabin1* (q 0.54), 1c. *dao* (q 0.48), 1d. *adh1b* (q 0.22), 1e. *camk2d* (p 0.25), 1f. *exoc1* (p 0.54), 1g. *whsc2* (p 0.78). **LG2:** 2a. *triti* (q 0.28), 2b. *uspl1* (q 0.75), 2c. *pan3* (q 0.76). **LG3:** 3a. *jhdmi1d* (p 1.00), 3b. *phf15* (q 0.10), 3c. *fgfr4* (q 0.26), 3d. *chd3* – cluster F (q 1.00). **LG4:** 4a. *sax6* (p 0.96), 4b. *ext2* (p 0.83), 4c. *acp2* (p 0.68), 4d. *rasgrp2* (p 0.23), 4e. *adams18* (q 0.14), 4f. *e2f4* (q 0.32), 4g. *mast2* (q 0.45), 4h. *pcsk9* (q 0.47), 4i. *dmap1* (q 0.48), 4j. *rybp* (q 0.87), 4k. *ppp4r2* (q 0.90), 4l. *ptprg* (q 0.96). **LG5:** 5a. *tram2* – cluster A (q 0.88), 5b. *ap2m1* (q 0.47), 5c. *myo6* (q 0.35), 5d. *cdc40* (q 0.25), 5e. *crim1* (p 0.26), 5f. *ttc27* (p 0.30), 5g. *cyp11b1* (p 0.41), 5h. *prepl* (p 0.77), 5i. *bre* (p 0.93). **LG6:** 6a. *atp6v1h* (q 0.55), 6b. *b4galt6* (q 0.54), 6c. *fign1* (p 0.24), 6d. *Xt7.1-THdA017108.3* (p 0.65), 6e. *mpp7* (p 0.68), 6f. *csmp1* (p 0.85). **LG7:** 7a. *nap2* (p 0.92), 7b. *gpr123* (p 0.59), 7c. *gatl* (p 0.46), 7d. *mfrn2* (q 0.14), 7e. *chl* (q 0.43), 7f. *rbcl* (q 0.43), 7g. *fur1* (q 0.78), 7h. *agmat* (q 0.95). **LG8:** 8a. *glpr2* – cluster 5B (p 0.92), 8b. *ENSXETG00000000667* (p 0.83), 8c. *dusp9* (q 0.04), 8d. *smc1a* (q 0.04), 8e. *nr6a1* (q 0.09), 8f. *rps6ka6* (q 0.36), 8g. *f9* (q 0.40), 8h. *zfp361i* (q 0.78), 8i. *me2d* (q 0.90), 8j. *rhbg* (q 0.92), 8k. *ubqln4* (q 0.93). **LG9:** 9a. *smarcal1* (q 1.00), 9b. *znf142* (q 0.98), 9c. *thr1* (q 0.74), 9d. *MGC145260* (p 0.58). **LG10:** 10a. *map2k4* (q 1.00), 10b. *src* (q 0.68), 10c. *rmf1* (p 0.28), 10d. *hoxb3* (p 0.51), 10e. *sp2* (p 0.53), 10f. *taf4* (p 0.75).

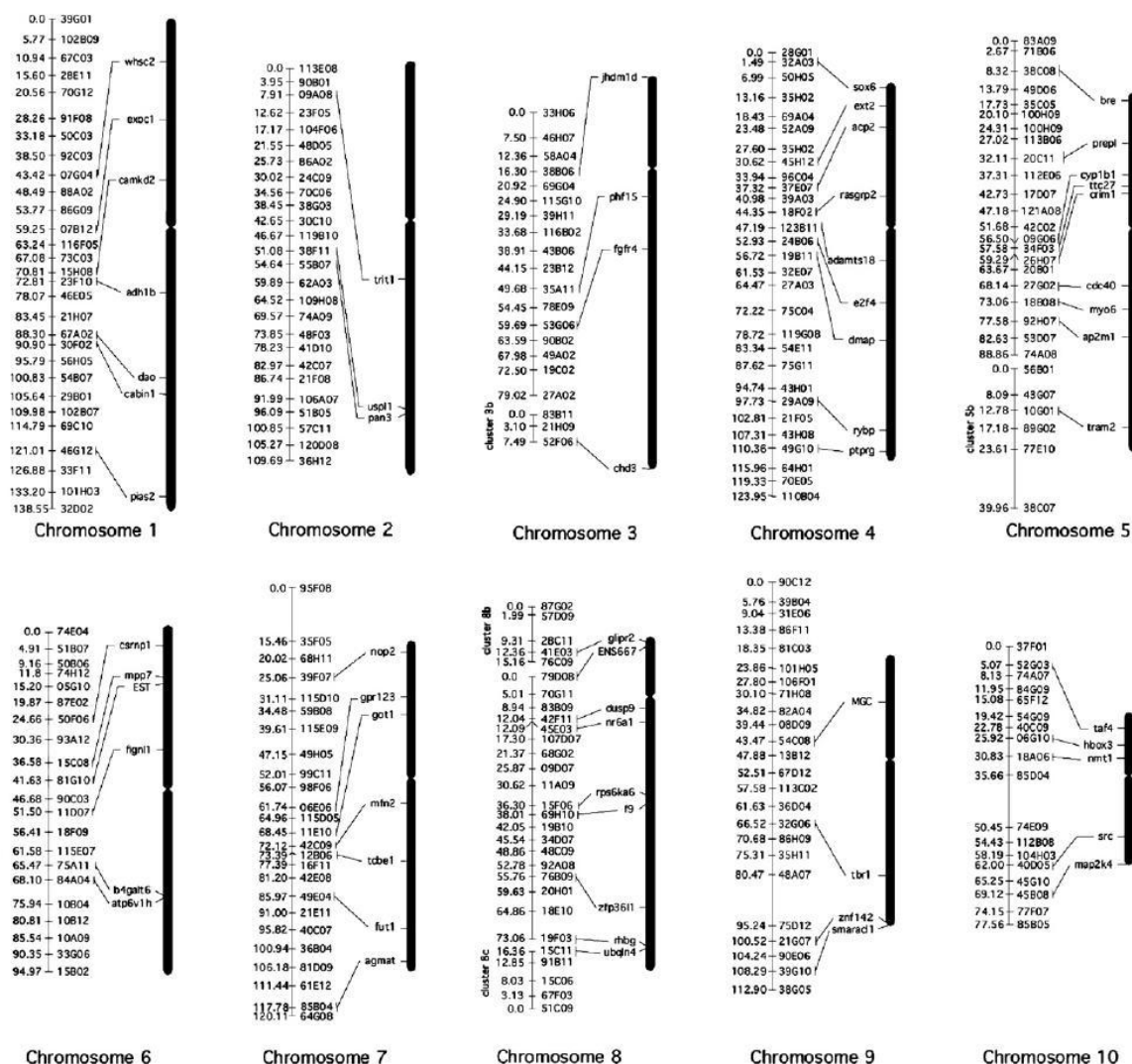


Fig. 3. Alignment of Physical and Genetic maps. Each of the ten chromosomes of *X. tropicalis* is depicted proportional to its size and centromere location. Locations of genes on each chromosome are shown based on *in situ* hybridization results. Representative markers and their centiMorgan positions from each of the linkage groups are shown alongside the chromosomes. The *in situ* mapped chromosomal genes are aligned to their closest linkage group marker. This linkage was established by comparing the location of the genetic markers and the location of the mapped genes within the same genomic scaffold.

on the p arm of Chr. 2 is substantially reduced relative to the regions represented on the map.

The genetic map and the genome assembly

We investigated the correspondence between genetic interval and genomic sequence by comparing the maximum genetic interval and genomic distance between markers on 74 of the largest 100 scaffolds; the remaining 26 scaffolds in this group were either represented by a single marker or were otherwise unsuitable. The ratio of genomic distance to genetic interval varied over 3 orders of magnitude, from approximately 4 kb/cM to over 3000 kb/cM; the average value is 581 kb/cM, comparable to the value of 625 kb/cM for zebrafish (Postlethwait et al., 1994). Interestingly, this value was not strongly correlated with distance from the centromere; scaffolds in proximity to the centromere included those with low ratios, as well as the expected higher kb/cM values, and some of the scaffolds with values over 1000 kb/cM were found far from the centromeric region (data not shown). Moreover, one-way ANOVA indicated that there were no

significant differences in the average map interval/scaffold ratios for all scaffolds among the individual linkage groups (data not shown). It should be noted, however, that the order of markers from a single scaffold on the genetic map often did not correspond to marker order in the genomic sequence within a scaffold, and that markers from a single scaffold were generally intermixed with markers from other scaffolds in the genetic map. Thus, for close-range analysis, the overall correspondence of marker order between the genetic map and the genomic sequence is relatively weak.

The striking short-range intermixing of markers from different scaffolds led us to ask whether this was a general feature of the map, or whether some linkage groups had a higher level of disjunct (i.e., intermixed) markers than others. As an indicator of "intermixing", we assessed each scaffold represented by more than one marker and identified the numbers of disjunct vs conjunct scaffolds for all scaffolds represented by only two markers (data not shown). The number of 2-marker scaffolds in each major linkage group (LG1–10) ranged from 5 to 27. For the 7 linkage groups with at least 10 2-marker scaffolds, the frequency of conjunct 2-marker scaffolds (i.e.,

scaffolds in which the representative markers are placed sequentially on the map, without intervening markers from other scaffolds) ranged from 26 to 36%; overall, 30% of 2-marker scaffolds were conjunct. We also compared the genetic distance between the representative markers for all of the 2-marker scaffolds for each linkage group (data not shown). For individual linkage groups with at least 10 2-marker scaffolds, between 50 and 72% were represented across a distance of 1 cM or less. Across all of the major linkage groups, over 60% of the 2-marker scaffolds are represented within a genetic distance of 1 cM. We tentatively conclude that the frequency of scaffold intermixing is relatively constant across the map and thus cannot be attributed to the specific properties of any individual linkage group.

Genomic representation

We estimated the minimum genomic representation for each linkage group by identifying the scaffolds from Assembly 4.1 that contained each of the mapped markers and determining the total size of scaffolds represented unambiguously in each linkage group. The minimum genomic representation varied between 29 and 164 MB across the 10 linkage groups.

Each linkage group included markers for which no single matching scaffold from Assembly 4.1 could be identified. These included markers that corresponded to sequences in multiple scaffolds, as well as markers that did not match any of the scaffolds from Assembly 4.1. Since each of the SSLP markers originated as a unique sequence from Assembly 2.0, these sequences have presumably been lost from the later genome assembly. The entire genetic map includes 126 sequences (4.3% of markers) with multiple corresponding scaffolds, and 122 sequences (4.2%) with no corresponding scaffold.

Scaffold matches for the remaining markers allowed us to assign individual scaffolds to linkage groups. For a scaffold to be assigned to a linkage group, 75% of markers included on the scaffold had to map to a single linkage group. The 45 scaffolds that did not meet this criteria were left “unassigned”, as listed in Table 1; they constitute just over 87 MB of the total genomic representation on the genetic map. The minimal genomic representation of the entire map totaled 1.129 GB, or 66% of the 1.7 GB represented in *X. tropicalis* genome assembly 4.1; the size of the assembled genome is similar to the genome sizes corresponding to the c-values obtained for *X. tropicalis* (1.47–1.88 GB; Gregory, 2006). Approximately half of the *X. tropicalis* genome assembly is represented in the first 272 scaffolds (Hellsten et al., 2010); all but 38 of these scaffolds are represented in the genetic map.

Scaffolds in the unassigned group included markers that mapped to multiple linkage groups. To evaluate the provenance of these scaffolds, we checked the associated marker sequences for their scaffold assignments in *X. tropicalis* genome assembly 5, a test genome assembly that was assembled using the ARACHNE compiler at the Stanford Genome Center. Out of the 45 “unassigned” Assembly 4.1 scaffolds, 30 included markers that were incorporated into different scaffolds in Assembly 5, suggesting that these 30 scaffolds from Assembly 4.1 may be chimeric. For 8 of the unassigned scaffolds, one or more SSLPs within the marker set could not be assigned to a single Assembly 5 scaffold, so they could not be evaluated. The remaining 7 scaffolds were cohesive, meaning that all markers within a given Assembly 4 scaffold mapped to one Assembly 5 scaffold. This result indicates that for these 7 cohesive scaffolds, there is a discrepancy between the genome assembly, which demonstrates that these sequences are in physical proximity to one another, and the genetic map, which places markers from a single scaffold into different linkage groups. We do not have an explanation for these findings, but caution those using this map for positional cloning studies to avoid using these scaffolds as a major source of linkage relationships. A list of unassigned scaffolds and their status with regard to assembly 5 is provided in Supplementary Table 3.

Alternative linkage analysis using MSTmap

In view of the discrepancies between the genetic map and genome assembly 4.1, we carried out a second linkage analysis using MSTmap, an alternative mapping program that establishes marker order using an algorithm based on the “minimal spanning tree” (MST) of a graph of the genotyping data (Wu et al., 2008). This alternative analysis required a significant recoding of the genotype data. In JoinMap 3.0, the F2 alleles were grouped in 5 classes based on the P1 genotype; each P1 genotype could generate 2–4 different F2 allelic combinations, depending on the number of alleles shared between the parents, for a total of 14 different types of F2 allelic combinations. In contrast, MSTmap recognizes only 3 types of F2 allelic combinations; thus, the transformation of these data into MSTmap format significantly reduced the complexity of the data set.

MSTmap analysis of the recoded genotypes yielded 96 linkage groups that included between 3 and 261 markers, 1 pair of linked markers, and 1 unlinked marker, incorporating 2326 markers in all. The total size of the map, representing the sum of the sizes of all linkage groups, is 1991.8 cM. Each MSTmap linkage group (lg) consisted of markers that could be found on a single linkage group from the primary Joinmap analysis (referred to as “LGs”); however, the markers within a single MSTmap lg cover large distances on the original Joinmap LGs, and the MSTmap lgs show considerable overlap on the corresponding Joinmap LGs. Interestingly, the MSTmap lgs tended to represent groupings of markers that displayed the same type of allelic combination in the original genotype data, suggesting that the MSTmap algorithm “sorts” markers by allelic combination as a byproduct of the computational process, despite the fact that specific information regarding the allelic combination for a given marker is lost during recoding. A spot-check comparison of the correspondence between the genome assembly and either of the two genetic maps for 38 scaffolds indicated that for 5 scaffolds, the markers from a given scaffold showed improved marker order (markers are more conjunct, i.e., less intermixing with markers from other scaffolds) in the MSTmap lgs. For 20 scaffolds, the markers were more conjunct in the Joinmap LGs, and for 13 scaffolds, the number of conjunct markers was identical in both linkage analyses (data not shown). Thus, in our view, while the MSTmap lgs confirm linkages shown by Joinmap, the MSTmap analysis does not produce the hoped-for improvement in marker order. We include the MSTmap linkage groups as a supplement (see Supplementary Table 4), as they may provide some alternative candidate linked markers.

We have not been able to combine MSTmap lgs that are included within the same Joinmap LG into a single linkage group in MSTmap. It is likely that the high level of parental heterozygosity for most of these markers precludes the detection of linkage using the MSTmap algorithm.

Discussion

We have generated an SSLP map of over 2800 markers, with an effective size of 1658 cM and a theoretical limit of resolution of 0.26 cM. This map is the third version that has been released through our website (<http://tropmap.biology.uh.edu>). A comparison of the size of the current 2886-marker map with the previous 1654-marker map suggests that 1658 cM may be near the maximum size of the genetic map, since the effective size of the previous map was 1636 cM. Thus, a 74% increase in the number of markers has yielded an increase of less than 2% in map size.

The minimal genomic representation of the entire map totaled to 1.129 GB, or 66% of the 1.7 GB represented in *X. tropicalis* genome assembly 4.1; the size of the assembled genome is similar to the genome sizes corresponding to the c-values obtained for *X. tropicalis* (1.47–1.88 GB; Gregory, 2006).

This map demonstrates that a bioinformatics-based strategy is an effective means of identifying SSLPs. The frequency of unique, usable SSLPs identified in our study was high relative to those of SSLPs identified via conventional CA-repeat screening of genomic libraries. Our recent analysis of 25% of the SSRs in the *X. tropicalis* genome (Xu et al., 2008) has shown that the frequency of all di-tri and tetranucleotide SSRs in *X. tropicalis* is close to 1/6 kb, much higher than previously suggested. Moreover, a bioinformatics strategy can be targeted to specific scaffolds to expand coverage of the genome.

A provisional AFLP map for *X. tropicalis* has been published (Kochan et al., 2003). This map includes 51 AFLP markers and 2 isozyme markers in 13 linkage groups. Since no sequence information is available for the AFLP markers, it is not possible to integrate this map with either our map or the genome assembly. Future mapping efforts may focus on SNP identification and mapping; since our strategy provides a correspondence between the genetic map with the genome sequence, it should be possible to integrate our map with future SNP mapping results.

This map shows two major discrepancies with the physical map. First, 8.5% of the markers on this map cannot be uniquely identified on a single scaffold from the current assembly (4.1; Hellsten et al., 2010); some markers are not represented on any scaffold in this assembly, while others show close matches to multiple scaffolds. Most of these are resolved in *X. tropicalis* genome assembly v6; comparisons to v6 indicate that only 38 markers cannot be unambiguously assigned to a single scaffold (Gilchrist and Zimmerman, pers. comm.). Second, there are discrepancies in short-range marker order, in that markers assigned to a single scaffold are intermixed along the genetic map with markers from other scaffolds; in many cases, these represent intermixing over short genetic distances. Most of these disjunct scaffolds are not resolved in later assemblies (Gilchrist and Zimmerman, pers. comm.). During our linkage analyses, we found that, although long-range marker position was robust, short-range marker position was often exquisitely sensitive to the presence of individual markers in flanking regions. Comparisons of genetic and 2 different sequence-based physical maps of the human genome suggested that most of the discrepancies could be attributed to errors in genome assembly (DeWan et al., 2002); however, this study was restricted to a limited set of markers that passed a more stringent test of marker position (likelihood ratio). The spatial heterogeneity of recombination frequencies along each chromosome may contribute to the apparent intermixing of markers within the genetic map. Increases in marker density on the genetic map should improve the short-range colinearity of these resources. Colinearity between the map and genome has already been increased by improvements to the genome assembly: the newly-released Assembly 7.1 shows greatly improved colinearity with the genetic map (Schmutz, Jenkins, and Rokhsar, personal communication). It should be noted, however, that the genetic map was incorporated in the construction of Assembly 7.1.

Graf (1989) generated a preliminary set of 8 linkage groups for *X. laevis* using 21 isozyme markers. We have attempted to compare these linkage groups to our map by finding the scaffold and linkage group assignments for the corresponding *X. tropicalis* genes. In many cases, the corresponding genes could not be unambiguously identified. We were, however, able to determine that the linkage relationships represented in Graf's linkage groups 1 and 6 are not conserved, although the 3 isozyme markers that constitute Graf's linkage group 7 are all found within our LG2 (data not shown). Further conclusions regarding the conservation of organization between the *X. tropicalis* and *X. laevis* genomes await sequencing and assembly of the *X. laevis* genome.

The striking absence of coverage on the p arm of Chr. 2 indicates that we have been unable to detect linkage either to or within this area. Although this omission could occur for a number of reasons, our results suggest that the level of polymorphism is lower in this region than in the regions of the genome that are represented on the genetic

map. Cytogenetic analyses of *X. tropicalis* lampbrush chromosomes (Penrad-Mobayed et al., 2009) should reveal any large-scale differences in crossover frequency for the p arm of Chr. 2. One explanation for a reduced frequency of polymorphism is "selective sweep", in which recent strong selection for an allele increases the frequency of alleles in regions flanking the locus at which selection occurs, via "genetic hitchhiking" (Dorit et al., 1995). Selective sweep is thought to underlie a dramatic reduction in polymorphisms over a 1-MB region in the maize genome, in comparison with its undomesticated relative teosinte, another subspecies of *Zea mays* (Tian et al., 2009). Analysis of single nucleotide polymorphism (SNP) distribution in the human genome has identified several genomic regions in which reduction of polymorphism may represent a selective sweep, including 2 areas located on the X chromosome (International HapMap Consortium, 2005). Studies of the Segregation Distorter (SD) complex in African populations of *Drosophila melanogaster* have demonstrated a recent selective sweep within the SD-mal variant, which carries a double inversion incorporating the region of the SD complex; on the SD-mal chromosome, recombination is suppressed across 14 MB (Presgraves et al., 2009). These authors point out that selfish meiotic drive complexes such as SD could become fixed in the absence of modifying loci or counteracting selection, leading to regional suppression of recombination. Although selective sweep alone is unlikely to explain the omission of the entire p arm from the genetic map, a localized region subject to selective sweep might disrupt our ability to detect linkage relationships over a larger area. Moreover, this relationship suggests that the p arm of chromosome 2 may harbor genes or gene complexes that could underlie previously undetected meiotic drive.

Suppression of recombination may occur in regions associated with sex determination. Although frogs use a ZW sex determination system, the Z and W chromosomes are morphologically indistinguishable (for review see Evans, 2008). A recent study indicates that 3 genes known to be sex-linked in the anuran *Rana rugosa* are located on *X. tropicalis* Chr. 8 (Tymowska Chr. 10) (Uno et al., 2008). The DM-W gene is essential for ovary development in *X. laevis* and may function in sex determination; bioinformatic comparisons, however, indicate that DM-W cannot be identified in the *X. tropicalis* genome (Yoshimoto et al., 2008). The locus responsible for sex determination in *X. tropicalis* remains unknown. Identification and sequencing of sex-linked AFLPs in *X. tropicalis* have been used to identify 9 scaffolds that show sex linkage (Olmstead et al., 2010); 4 of these are represented in LG7, and the remainder is not represented on the map.

This map will be a significant resource for genetic and genomic studies of *X. tropicalis*. Preliminary versions of this map have already been used to map genes identified by mutation (Abu-Dea et al., 2009) and track the movement of transposable elements (Yergeau et al., 2010). The genetic map, together with the marker-scaffold correspondence, has also contributed to the evaluation of predicted syntenic relationships and long-range assembly of the *X. tropicalis* genome (Hellsten et al., 2010). Given the map size, resolution, marker density, and relative ease of use, the map should be valuable for the integration of genetics with genomic or cytogenetic studies.

Supplementary materials related to this article can be found online at doi:10.1016/j.ydbio.2011.03.022.

Acknowledgments

Identification of SSLPs and construction of the genetic map were supported by NIH RO1 HD046661 to DEW, AKS, and SES. The physical mapping studies were supported by grant 304/07/P204 from the Grant Agency of the Czech Republic and by grant MSM0021620858 from the Ministry of Education, Youth and Sports of the Czech Republic to V. K. and J. M. D.L.S. was supported by the Wellcome Trust. The F1 animals from the initial P1 cross were provided by Rob Grainger. We wish to thank Vijay Dontharaju, Joseph Gall, Richard Gibbs, Mike Gilchrist, Rob Grainger, Richard Harland, Jeremy Jenkins,

Steve Klein, Chih-Hong Lou, Dan Rokhsar, Jeremy Schmutz, Yi-Er Shieh, Akif Uzman, George Weinstock, Diane Wiernasz, Jim Wright, and Lyle Zimmerman for suggestions, support, and assistance.

References

- Abu-Daya, A., Sater, A.K., Wells, D.E., Mohun, T.J., Zimmerman, L.B., 2009. Absence of heartbeat in the *Xenopus tropicalis* mutation *muzak* is caused by a nonsense mutation in cardiac myosin *myh6*. *Dev. Biol.* 336, 20–29.
- Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580.
- DeWan, A.T., Parrado, A.R., Matisse, T.C., Leal, S.M., 2002. The map problem: a comparison of genetic and sequence-based physical maps. *Am. J. Hum. Genet.* 70, 101–107.
- Dorit, R.L., Akashi, H., Gilbert, W., 1995. Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science* 268, 1183–1185.
- Evans, B.J., 2008. Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). *Front. Biosci.* 13, 4687–4706.
- Goda, T., Abu-Daya, A., Carruthers, S., Clark, M.D., Stemple, D.L., Zimmerman, L.B., 2006. Genetic screens for mutations affecting development of *Xenopus tropicalis*. *PLoS Genet.* 2, e91.
- Graf, J.-D., 1989. Genetic mapping in *Xenopus laevis*: eight linkage groups established. *Genetics* 123, 389–398.
- Grammer, T.C., Khokha, M.K., Lane, M.A., Lam, K., Harland, R.M., 2005. Identification of mutants in inbred *Xenopus tropicalis*. *Mech. Dev.* 122, 263–272.
- Gregory, T.R., 2006. Animal Genome Size Database. <http://www.genomesize.com>.
- Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., Blitz, I., Blumberg, B., Dichmann, D., Dubchak, I., Fletcher, R., Gerhard, D., Goodstein, D., Graves, T., Grigoriev, I., Grimwood, J., Kawashima, T., Lindquist, E., Mead, P., Mitros, T., Ota, Y., Poliakov, A., Pollet, N., Robert, J., Salamov, A., Sater, A.K., Schmutz, J., Terry, A., Vize, P., Warren, W., Wells, D., Wills, A., Zimmerman, L.B., Grainger, R.M., Grammer, T., Khokha, M., Richardson, P.M., Rokhsar, D.S., 2010. The genome of the western clawed frog *Xenopus tropicalis*. *Science* 328, 633–636.
- International HapMap Consortium, 2005. A haplotype map of the human genome. *Nature* 437, 1299–1320.
- Khokha, M.K., Krylov, V., Reilly, M.J., Gall, J.G., Bhattacharya, D., Cheung, C.Y., Kaufman, S., Lam, D.K., Macha, J., Ngo, C., Prakash, N., Schmidt, P., Tlapakova, T., Trivedi, T., Tumova, L., Abu-Daya, A., Geach, T., Vendrell, E., Ironfield, H., Sinzelle, L., Sater, A.K., Wells, D.E., Harland, R.M., Zimmerman, L.B., 2009. Rapid gynogenetic mapping of *Xenopus tropicalis* mutations to chromosomes. *Dev. Dyn.* 238, 1398–46.
- Knapik, E.W., Goodman, A., Atkinson, O.S., Roberts, C.T., Shiozawa, M., Sim, C.U., Weksler-Zangen, S., Trolliet, M.R., Futrell, C., Innes, B.A., Koike, G., McLaughlin, M.G., Pierre, L., Simon, J.S., Vilallonga, E., Roy, M., Chiang, P.W., Fishman, M.C., Driever, W., Jacob, H.J., 1996. A reference cross DNA panel for zebrafish (*Danio rerio*) anchored with simple sequence length polymorphisms. *Development* 123, 451–460.
- Kochan, K.J., Wright, D.A., Schroeder, L.J., Shen, J., Morizot, D.C., 2003. Genetic linkage maps of the West African clawed frog *Xenopus tropicalis*. *Dev. Dyn.* 226, 99–102.
- Krylov, V., Tlapakova, T., Macha, J., 2007. Localization of the single copy gene *Mdh2* on *Xenopus tropicalis* chromosomes by FISH-TSA. *Cytogenet. Genome Res.* 116, 110–112.
- Kuittinen, H., de Haan, A.A., Vogl, C., Oikarinen, S., Leppälä, J., Koch, M., Mitchell-Olds, T., Langley, C.H., Savolainen, O., 2004. Comparing the linkage maps of the close relatives *Arabidopsis thaliana* and *A. thaliana*. *Genetics* 168, 1575–1584.
- Mácha, J., Teichmanová, R., Sater, A.K., Wells, D., Tlapáková, T., Zimmerman, L., Krylov, V., manuscript in preparation. Evolution of mammalian X chromosome and *Xenopus tropicalis* genome.
- McGinnis, S., Madden, T.L., 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 32, W20–W25 (Web Server issue).
- Noramly, S., Zimmerman, L., Cox, A., Aloise, R., Fisher, M., Grainger, R.M., 2005. A gynogenetic screen to isolate naturally occurring recessive mutations in *Xenopus tropicalis*. *Mech. Dev.* 122, 273–287.
- Olmstead, A.W., Lindberg-Livingston, A., Degitz, S.J., 2010. Genotyping sex in the amphibian, *Xenopus (Silurana) tropicalis*, for endocrine disruptor bioassays. *Aquat. Toxicol.* 98, 60–66. online 25 January 2010.
- Penrad-Mobayed, M., El Jamil, A., Kanhoush, R., Perrin, C., 2009. Working map of the lampbrush chromosomes of *Xenopus tropicalis*: a new tool for cytogenetic analyses. *Dev. Dyn.* 238, 1492–1501.
- Postlethwait, J.H., Johnson, S., Midson, C.N., Talbot, W.S., Gates, M., Ballinger, E.W., Africa, D., Andrews, R., Carl, T., Eisen, J.S., Horne, S., Kimmel, C.B., Hutchinson, M., Johnson, M., Rodriguez, A., 1994. A genetic map for the zebrafish. *Science* 264, 699–703.
- Presgraves, D.C., Gerard, P.R., Cherukuri, A., Lyttle, T.W., 2009. Large-Scale Selective Sweep among Segregation Distorter Chromosomes in African Populations of *Drosophila melanogaster*. *PLoS Genet.* 5 (5), e1000463.
- Rosen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132, 365–386.
- Song, Q.J., Marek, L.F., Shoemaker, R.C., Lark, K.G., Concibido, V.C., Delannay, X., Specht, J.E., Cregan, P.B., 2004. A new integrated genetic linkage map of the soybean. *Theor. Appl. Genet.* 109, 122–128.
- Sparrow, D.B., Latinkic, B., Mohun, T.J., 2000. A simplified method of generating transgenic *Xenopus*. *Nucleic Acids Res.* 28, E12.
- Stam, P., 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J.* 3, 739–744.
- Strachan, T., Read, A., 2003. Human Molecular Genetics, Third Edition. Garland Science, New York.
- Tian, F., Stevens, N.M., Buckler IV, E.S., 2009. Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *PNAS USA* 106, 9979–9986.
- Uno, Y., Nishida, C., Yoshimoto, S., Ito, M., Oshima, Y., Yokoyama, S., Nakamura, M., Matsuda, Y., 2008. Diversity in the origins of sex chromosomes in anurans inferred from comparative mapping of sexual differentiation genes for three species of the Raninae and Xenopodinae. *Chromosome Res.* 16, 999–1011.
- Van Ooijen, J.W., Voorrips, R.E., 2001. JoinMap® Version 3.0. Software for the calculation of genetic linkage map. Plant Research International, Wageningen, Netherlands.
- Wu, Y., Bhat, P.R., Close, T.J., Lonardi, S., 2008. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet.* 4, e1000212.
- Xu, Z., Gutierrez, L., Hitchens, M., Scherer, S., Sater, A.K., Wells, D.E., 2008. Distribution of polymorphic and non-polymorphic microsatellite repeats in *Xenopus tropicalis*. *Bioinf. Biol. Insights* 2, 157–169.
- Yergeau, D.A., Kelley, C.M., Kuliye, E., Zhu, H., Sater, A.K., Wells, D.E., Mead, P.E., 2010. Remobilization of Tol2 transposons in *Xenopus tropicalis*. *BMC Dev. Biol.* 10, 11.
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., Matsuda, Y., Takamatsu, N., Shiba, T., Ito, M., 2008. A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc. Nat. Acad. Sci. U.S.A.* 105, 2469–2474.

Preparation of *Xenopus tropicalis* whole chromosome painting probes using laser microdissection and reconstruction of *X. laevis* tetraploid karyotype by Zoo-FISH

Vladimir Krylov · Svatava Kubickova ·
Jiri Rubes · Jaroslav Macha · Tereza Tlapakova ·
Eva Seifertova · Natasa Sebkova

Received: 9 December 2009 / Accepted: 16 March 2010 / Published online: 14 April 2010
© Springer Science+Business Media B.V. 2010

Abstract Laser microdissection was used for the preparation of whole chromosome painting probes in *Silurana (Xenopus) tropicalis*. Subsequent cross-species fluorescence in situ hybridization (Zoo-FISH) on its tetraploid relative *Xenopus laevis* revealed persistence of chromosomal quartets even after 50–65 million years of separate evolution. Their arrangement is in a partial concordance with previous experiments based on similarity of a high-resolution replication banding pattern. Further support for an allotetraploid origin of *X. laevis* was given by hybridization with a probe derived from the smallest *X. tropicalis* chromosome (Xt10). Here, pericentric areas of both arms of XI 14 and 18 were stained, indicating intrachromosomal rearrangements. The positions of signals were not in agreement with the

chromosomal quartets revealed by painting probes Xt 8 and 9 (XI 11+14 and XI 15+18, respectively). This suggests that both *X. tropicalis* chromosomes underwent non-reciprocal translocation of Xt10 separately in at least two different ancient ancestors. In addition, the observed translocation events could explain the origin of individuals with 18 chromosomes in diploid karyotypes, probably extinct after the genesis of the allotetraploid *X. laevis* ($2n=36$).

Keywords *Xenopus tropicalis* · *Xenopus laevis* · whole chromosome painting probe · Zoo-FISH · microdissection

Abbreviations

DAPI	4',6-diamidino-2-phenylindole
DOP-PCR	degenerate oligonucleotide primed polymerase chain reaction
FISH	fluorescence in situ hybridization
MYA	million years ago
NOR	nucleolus organizer region
SSR	simple sequence repeat
TSA	tyramide signal amplification
WCP	whole chromosome painting
WGA	whole genome amplification
XI	<i>Xenopus laevis</i> chromosome
Xt	<i>Xenopus tropicalis</i> chromosome
Zoo-FISH	cross-species fluorescence in situ hybridization

Responsible Editor: Walther Traut.

V. Krylov (✉) · J. Macha · T. Tlapakova · E. Seifertova ·
N. Sebkova
Faculty of Science, Charles University in Prague,
Vinicna 7,
128 44 Prague 2, Czech Republic
e-mail: vkrylov@natur.cuni.cz

S. Kubickova · J. Rubes
Veterinary Research Institute,
Hudecova 70,
621 00 Brno, Czech Republic

Introduction

Xenopus tropicalis represents an amenable model organism for genetic and developmental studies. Its diploid genome containing ten pairs of easily distinguishable chromosomes (Tymowska 1973) was sequenced and arranged in scaffolds (Klein et al. 2002). At present, a provisional linkage map based on SSR (simple sequence repeat) polymorphic markers is available (<http://tropmap.biology.uh.edu>). The cytogenetics of *X. tropicalis* was established through the introduction of fluorescence in situ hybridization coupled with a tyramide signal amplification step (FISH-TSA) (Krylov et al. 2007). The mentioned technique was successfully employed for the identification of centromeric positions for all ten linkage groups. This information enabled the assignment of induced mutations to specific chromosomes in gynogenetic offspring (Khokha et al. 2009). Besides indisputable advantages of this model organism in the mentioned research fields, *X. tropicalis* and its tetraploid relative *Xenopus laevis* constitute pivotal creatures from the gene duplication point of view (Pollet and Mazabraud 2006). Amphibians in general are a connecting link between lower vertebrates, such as fishes, and terrestrial tetrapods including human. Despite their interesting taxonomic position, there is an almost complete lack of tools for the study of their chromosomal evolution. Until now, only chromosome banding and comparison of linkage maps were used (Hotz et al. 1997; Schmid and Steinlein 2001; Smith and Voss 2006).

Whole chromosome painting (WCP) coupled with cross-species fluorescence in situ hybridization (Zoo-FISH) can overcome this limitation. These techniques are based on the isolation of identical chromosomes via flow sorting or microdissection from metaphase spreads. An appropriate probe is prepared by nonspecific amplification using degenerate primers and labeling with a fluorophore-coupled or haptenized nucleotide. Subsequent hybridization with chromosomes of the same species (painting FISH) results in WCP. On the other hand, the term Zoo-FISH refers to using the probe on evolutionarily more or less related counterparts. These techniques are well-established in mammals and marsupials (Chowdhary and Raudsepp 2001), birds (Shetty et al. 1999; Chowdhary and Raudsepp 2000; Guttenbach et al. 2003; Griffin et al. 2008), fishes (Rab et al. 2008) and some nonverte-

brate taxa (Fukova et al. 2007; Vitkova et al. 2007; Teruel et al. 2009).

In this article, we describe the preparation of *X. tropicalis* whole chromosome painting probes using laser microdissection and show the reconstruction of allotetraploid origin of *X. laevis* by Zoo-FISH.

Materials and methods

Preparation of *X. tropicalis* and *X. laevis* metaphase spreads

X. tropicalis metaphase spreads were prepared from euploid testicular primary cell culture as described in Khokha et al. (2009) with minor changes. Briefly, the trypsinization step was followed by a hypotonic shock in 38-mM KCl for 5 min. After fixation, cell suspension was stored overnight at -20°C . For laser microdissection, cells were dropped on a polyethylene naphthalene membrane (P.A.L.M. GmbH, Bernried, Germany) attached to a thin glass slide, dried and stained with 3% Giemsa for 10 min. Painting FISH experiments were done with chromosomes spread onto clean microscopic slides as described in Courtet et al. (2001).

X. laevis chromosome samples were prepared from splenocytes of an adult female as described in Krylov et al. (2003) with minor changes. Briefly, 4.5 h before decapitation, the frog was injected with 10 $\mu\text{g/g}$ (w/w) of colchicine (2 mg/ml) into the lymphatic sac. The spleen was disrupted with two needles in 2 ml of diluted PBS (two volumes of PBS and one volume of water). The suspension was filtered through a polyamide sieve with 60- μm meshes. Hypotonization, fixation and the following treatment of slides were done as described above for *X. tropicalis*.

Laser microdissection of *X. tropicalis* chromosomes and preparation of WCPs

Chromosome dissection and collection was carried out using a PALM MicroLaser system (P.A.L.M. GmbH, Bernried, Germany) coupled with an inverted microscope (Olympus) under an oil immersion objective (100 \times magnification). About 15 copies of the desired chromosome were microdissected and catapulted by a single laser pulse directly into the cap of a PCR tube containing 4- μl PCR oil.

Whole chromosome painting probes were prepared in two ways: by two rounds of DOP-PCR (degenerate oligonucleotide primed PCR), with Dig-11-dUTP (Roche Diagnostic GmbH, Germany) as described in Kubickova et al. (2002) or by GenomePlex Single Cell Whole Genome Amplification Kit (WGA4) (Sigma-Aldrich Co., USA), according to the manufacturer's manual followed by purification on a Gel extraction kit column (Qiagen GmbH, Germany). Labeling of 20 ng of the amplification product was performed with the GenomePlex DNA Reamplification kit (WGA3) (Sigma-Aldrich Co., USA) and Dig-11-dUTP (Roche Diagnostics GmbH, Germany) with minor changes. The 10-mM stock solution of dNTPs from the kit was replaced by the following mixture (10 mM of dATP, dCTP, dGTP and 6.5 mM dTTP). The volume of this solution was reduced to a half (1.5 μ l instead of 3.0 μ l per 75 μ l reaction). The volume of 1-mM Dig-11-dUTP was 2 μ l per 75 μ l reaction. Cycling parameters were set according to the manufacturer's manual. After reamplification, the probe was purified on a Gel extraction kit column (Qiagen GmbH, Germany).

Painting FISH on *X. tropicalis* chromosomes

Twenty two microliters of the hybridization mixture containing 50% formamide, 2 \times SSC, 10% dextran sulfate, 5- μ g *X. tropicalis* competitor DNA (Applied Genetics Laboratories, Melbourne, USA) and 2 μ l of probe (approx. 500 ng) was denatured for 10 min at 72°C and then reannealed at 37°C for 80 min. During this time, slides with chromosomal spreads were denatured in 70% formamide, 2 \times SSC (pH 7.0) for 2 min, immediately dehydrated in methanol series (70, 90 and 100%), 3 min each with agitation and dried. The hybridization mixture was placed on a slide, covered with a 22 \times 22 mm coverslip and hybridized overnight in a wet chamber at 37°C. Posthybridization washing was done twice in 50% formamide, 2 \times SSC (pH 7.0) at 42°C, three times in 2 \times SSC at RT and once in 1 \times TNT at RT, 5 min each. Slides were then blocked with TNB buffer (100-mM Tris-HCl and 150-mM NaCl), 0.5% blocking reagent (Boehringer Mannheim GmbH, Germany) and pH 7.5 for 30 min. The detection of the probe was done with antidigoxigenin-tetramethylrhodamine (Roche Diagnostics GmbH, Germany). After washing with TNT buffer, slides were mounted in a mounting medium

with DAPI (4',6-diamidino-2-phenylindole) (Intimex, Czech Republic). Visualization of probe signals was performed under an Olympus BX40 microscope. Pictures were taken with a SONY Exwave HAD black and white camera and images were processed, colored and merged with the ACC Program, v. 5.0 (SOFO, Czech Republic).

Zoo-FISH on *X. laevis* chromosomes

The protocol for Zoo-FISH experiments was similar as for the painting FISH with minor changes. The volume of the probe was 3 μ l (approx. 750 ng), and hybridization time was prolonged up to 72 h. Post-hybridization washing with 50% formamide in 2 \times SSC was done at 38°C.

Chromosome identification and nomenclature

Chromosomes were identified by means of their relative length and p/q arms ratio according to Tymowska (1973) for *X. tropicalis* and Tymowska and Kobel (1972) for *X. laevis*. The numbering of *X. tropicalis* chromosomes follows the new ordering described in Khokha et al. (2009).

Results

Painting FISH on *X. tropicalis* chromosomes

The efficiency of DNA amplification was tested on chromosome 9 (Xt 9), one of the two containing a secondary constriction. After laser microdissection, amplification and labeling were done by two rounds of DOP-PCR with Dig-11-dUTP as described in Kubickova et al. (2002). Hybridization of the probe resulted in the labeling of a limited area including the secondary constriction on the q arm of Xt 9 (Fig. 1c). Interestingly, the signals in interphase nuclei clearly correspond to the two small nuclear bodies observed with DAPI staining (Fig. 1a and b). Since the painting of Xt 9 was only partial, we performed primary amplification of the same chromosome by GenomePlex Single Cell Whole Genome Amplification Kit (WGA4). This method utilizes random fragmentation of genomic DNA and amplification of small fragments (approx. 400 bp) by annealing primers with adaptors containing universal priming sites (Gribble

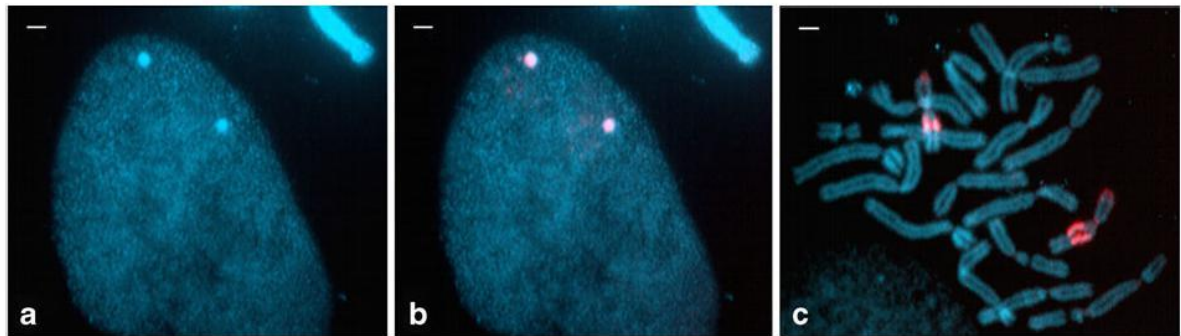


Fig. 1 Painting FISH with the Xt 9 probe prepared by two rounds of DOP-PCR. **a** A pair of nuclear bodies visible with DAPI staining. **b** FISH signals correspond to DAPI-stained nuclear bodies. **c** Painting FISH of the Xt9 probe on an *X.*

tropicalis metaphase spread. A strong signal is detected in the limited region including the secondary constriction of the appropriate chromosomal pair. Scale bar represents 2 μ m

et al. 2004). The following reamplification by GenomePlex WGA reamplification kit (WGA3) with Dig-11-dUTP resulted in a strong and specific painting probe covering approximately 80% of the particular chromosomal pair (Fig. 2g). Although the prepared probe was still not perfect, we tried to use the same system for the preparation of the remaining painting probes. As a result, seven probes for Xt 1, Xt 3, Xt 5, Xt 6, Xt 7, Xt 8 and Xt 10 hybridized specifically and stained whole chromosomes (Fig. 2). Unfortunately, we were not able to get usable probes for Xt 2 and Xt 4 where labeling was not specific. This might have been caused by accidental mixing of these two chromosomes in the microdissection step as their morphology is very similar.

Zoo-FISH on *X. laevis* chromosomes

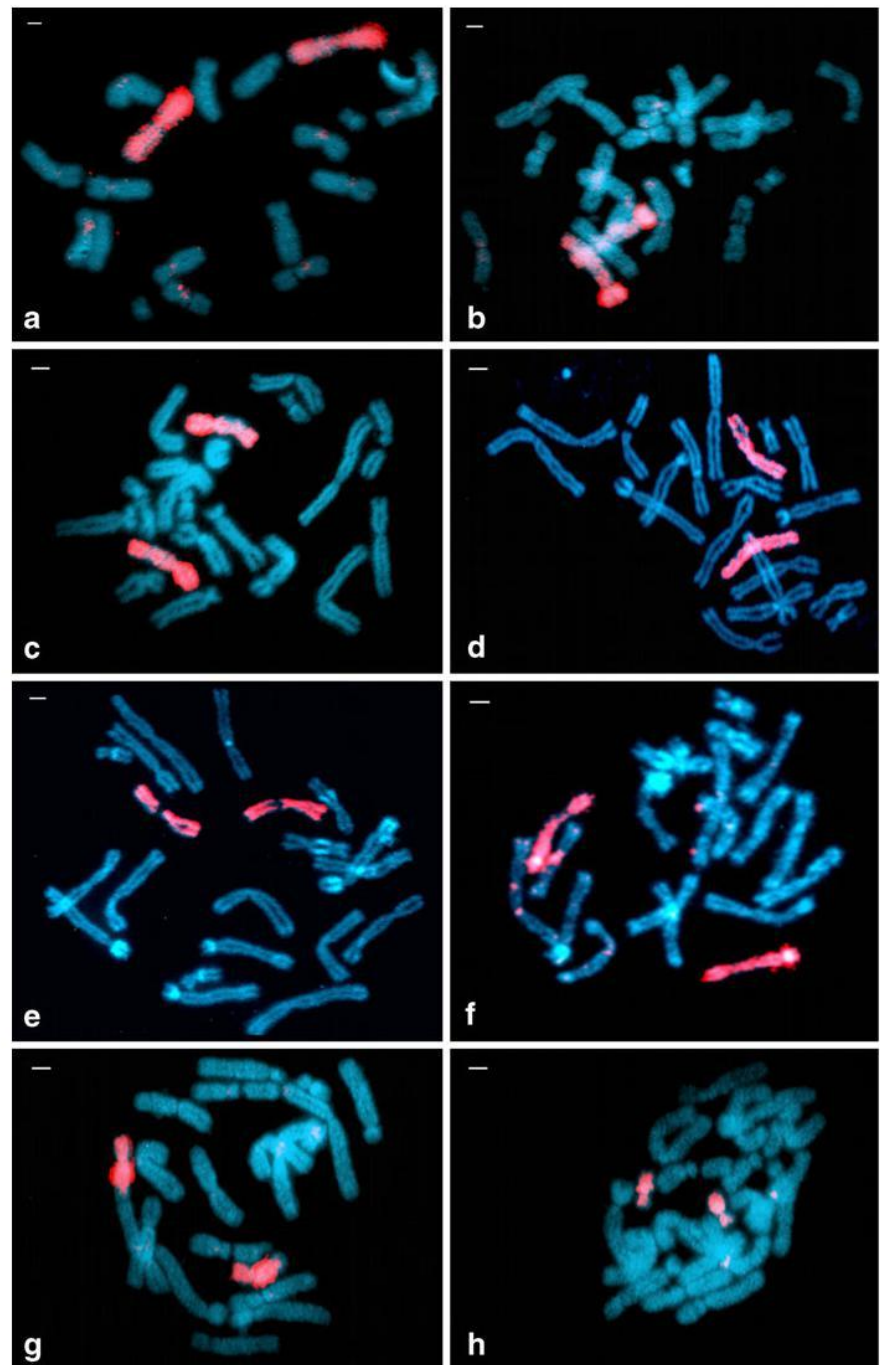
After the confirmation of probe specificity, we performed cross-hybridization (Zoo-FISH) experiments on *X. laevis* metaphase spreads. Four of eight probes labeled whole chromosomal quartets: Xt 1 (XI 1+2), Xt 5 (XI 13+17), Xt 6 (XI 6+9) and Xt 8 (XI 11+14) (Fig. 3a, c, d and f). The Xt 3 painting probe labeled the q arm of XI 12 and whole XI 16 (Fig. 3b). The probe derived from Xt 9 stained approximately one half of the q arms of XI 15+18 (Fig. 3g). A more complicated situation arose with the two remaining chromosomes (Xt 7 and Xt 10). The painting probe derived from Xt 7 labeled the whole XI 7+10 quartet, the satellite of XI 12 and a substantial part of the p arm of only one chromosome XI 13. In addition, fainter dispersed signals were observed on most other chromosomes (Fig. 3e). The probe from the smallest

X. tropicalis chromosome (Xt 10) stained the pericentric area on both arms of XI 14+18. Several minor signals were scattered largely on the p arms of some other acrocentric chromosomes (Fig. 3h). A schematized reconstruction of the *X. laevis* tetraploid karyotype based on Zoo-FISH experiments is depicted in Fig. 4.

Discussion

Genera *Silurana* and *Xenopus* diverged from the common ancestor approximately 50–65 million years ago (MYA) (Evans et al. 2004; Hellsten et al. 2007). Subsequent polyploidization events resulted in genomes of multiples of 20 chromosomes in the genus *Silurana* (*Xenopus epitropicalis*, $2n=40$) and of 18 chromosomes in the genus *Xenopus* (*X. laevis*, $2n=36$). So far, *Silurana* (*Xenopus*) *tropicalis* represents the only diploid species known in the group of clawed frogs ($2n=20$) (Evans 2008). The hypothetical model concerning the origin of $2n=18$ individuals was described in Schmid and Steinlein (1991). The karyotype of *X. tropicalis* contains one chromosomal pair Xt 10, not present in any other *Xenopus* species. The authors assume that in the African Miocene there existed several species with $2n=20$ and that in two of them, the chromosomal pair 10 could have been translocated by non-reciprocal rearrangements onto two different chromosomes. As a result, these individuals would contain 18 chromosomes in the diploid set. Our findings strongly support this theory since the painting probe derived from Xt 10 labeled

Fig. 2 Painting FISH with *X. tropicalis* whole chromosome painting (WCP) probes prepared by GenomePlex WGA4 amplification and WGA3 reamplification and labeling. Staining of the appropriate chromosomal pair using WCPs derived from **a** Xt 1, **b** Xt 3, **c** Xt 5, **d** Xt 6, **e** Xt 7, **f** Xt 8, **g** Xt 9, **h** Xt 10. Scale bar represents 2 μ m



small distinct pericentric areas on both arms of XI 14 and 18, indicating visible intrachromosomal rearrangements. In addition, minor signals were detected largely on the p arms of some other acrocentric

chromosomes, suggesting further interchromosomal translocations (Fig. 3h). Zoo-FISH experiments with Xt 8 and 9 probes revealed staining of chromosomal quartets XI 11+14 and XI 15+18, respectively

Fig. 3 Zoo-FISH on *X. laevis* metaphase spreads using *X. tropicalis* WCPs prepared by the GenomePlex WGA4/WGA3 protocol. **a** The Xt 1 probe stained the chromosomal quartet X1 1+2. **b** Hybridization with the Xt 3 probe gave staining of the q arms of X1 12 and the whole chromosome pair X1 16. **c** The Xt 5 probe stained the chromosomal quartet X1 13+17. **d** Zoo-FISH with the Xt 6 probe revealed a chromosomal quartet formed by X1 6 and 9. **e** Hybridization with the Xt 7 probe stained the chromosomal quartets X1 7 and 10. Strong signals were detected on X1 12 satellites and on the substantial part of the p arm of one chromosome from the X1 13 pair. **f** Hybridization with the Xt 8 probe resulted in staining of the chromosomal quartets X1 11 and 14. **g** Zoo-FISH with the Xt 9 probe revealed labeling of approximately one half of the q arms of the chromosomal quartet X1 15+18. **h** Hybridization with the probe derived from the smallest *X. tropicalis* chromosome (Xt10) gave pericentric staining on both arms of X1 14 and 18. Minor signals were scattered largely on the p arms of other acrocentric chromosomes. *Arrows* indicate chromosomal quartets. *Arrowheads* point to chromosomes with pericentric staining using the Xt 10 painting probe. *Asterisks* indicate probable visualization of the active NOR and polymorphic minor rDNA loci revealed by the Xt 7 painting probe. Scale bar represents 2 μ m

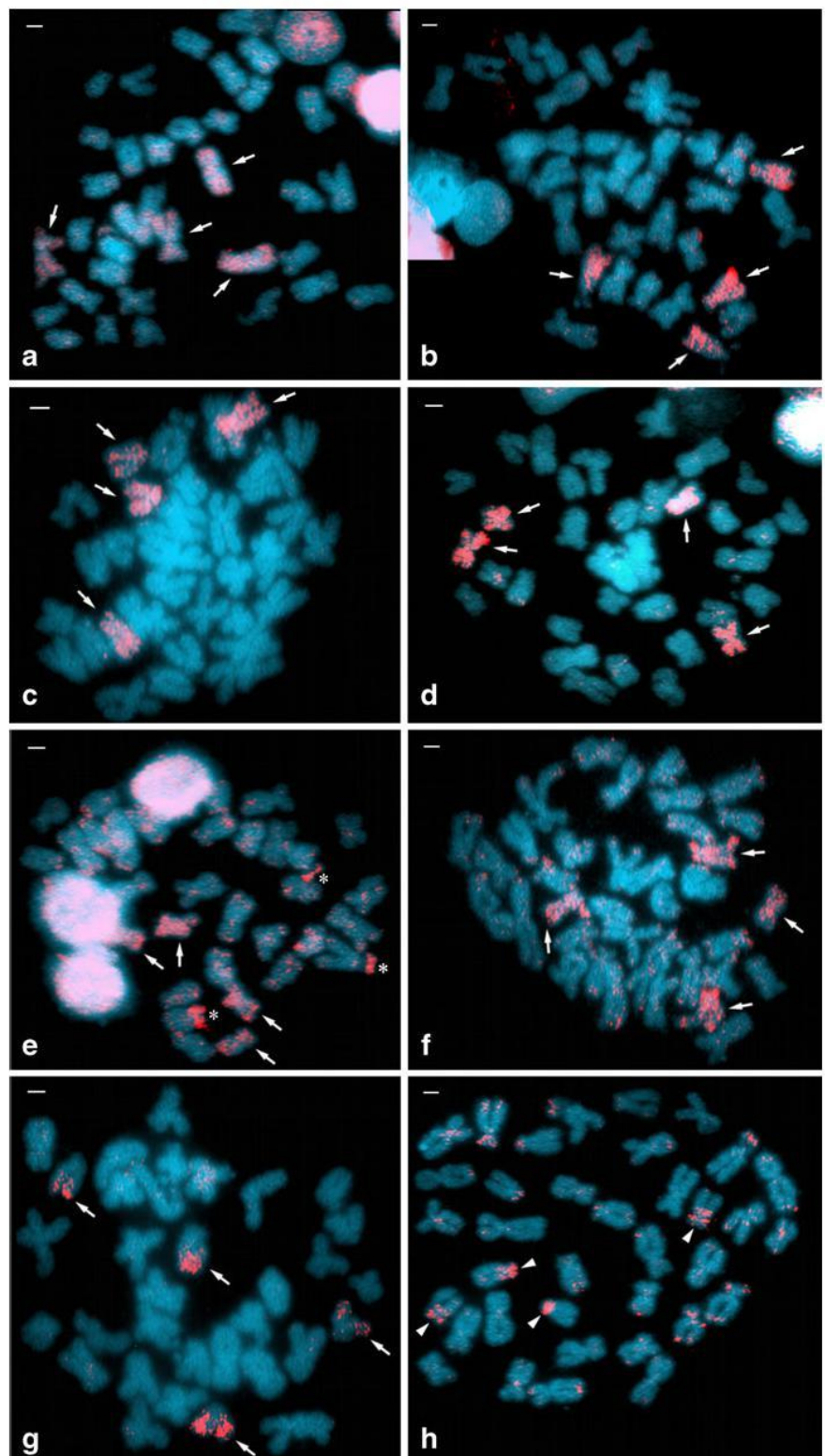
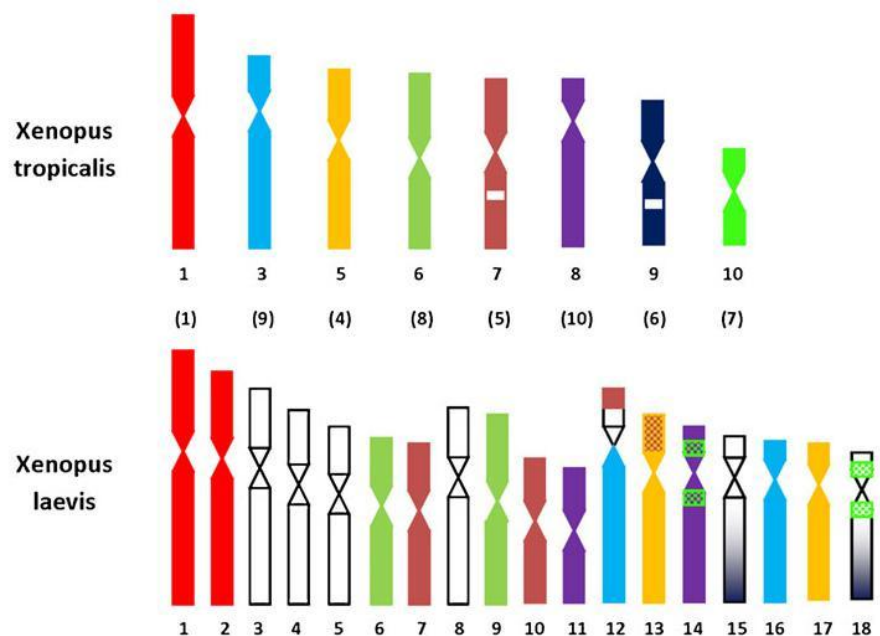


Fig. 4 Schematized reconstruction of the allotetraploid *X. laevis* karyotype based on data from Zoo-FISH experiments using *X. tropicalis* whole chromosome painting probes. Chromosome numbering is in accordance with Khokha et al. (2009) for *X. tropicalis* and Tymowska and Kobel (1972) for *X. laevis*. The numbers in brackets under *X. tropicalis* chromosomes refer to the old system described in Tymowska (1973)



(Fig. 3f and g). This observation supports an allotetraploid origin of *X. laevis*. The first ancestor underwent an Xt 10 translocation to Xt 8, in the second case, the recipient chromosome was Xt 9.

Hybridization of the Xt 9 painting probe prepared by GenomePlex kits on *X. tropicalis* metaphase spreads resulted in staining of approximately 80% of the appropriate chromosomal pair (Fig. 2g). In addition, both the GenomePlex as well as DOP-PCR probes revealed a strong signal in the secondary constriction and two small round structures in interphase nuclei which corresponded to DAPI-stained nuclear bodies (Figs. 1 and 2g). A possible explanation is the presence of the nucleolus organizer region (NOR). However, an active NOR was previously localized by silver staining to the secondary constriction in Xt 7 (Tymowska and Fischberg 1982) and the satellite of XI 12 in *X. laevis* (Schmid and Steinlein 1991). Hybridization of the Xt 7 painting probe with *X. laevis* chromosomes revealed the staining of the chromosomal quartet XI 7+10, the XI 12 satellite and a substantial part of the p arm of only one chromosome from the XI 13 pair (Fig. 3e). This observation indirectly corroborates the positions of active NORs in both species. Fainter dispersed signals on most other chromosomes probably represent additional rDNA loci. The presence of the

redundant signal on the single XI 13 could be explained by individual polymorphism of the minor rDNA loci as described previously in salmonid fishes (Fujiwara et al. 1998) or the Atlantic halibut (Ocalewicz et al. 2008). Since the Zoo-FISH experiments with Xt 9 didn't show any distinct regions in metaphase spreads or in interphase nuclei, we conclude that the secondary constriction present on this chromosome was lost by deletion. As a result, only XI 12 bears the constriction from Xt 7. An analogous situation was observed in the tetraploid and octaploid South American anurans *O. americanus* and *Ceratophrys ornate* (Schmid et al. 1985).

High abundance of repetitive sequences in Xt 9 could result in their preferential strong amplification during preparation of the painting probe. Consequently, the probe prepared by two rounds of DOP-PCR reacts only with the secondary constriction and its adjacent region (Fig. 1c). Even the superior GenomePlex probe gives partial Xt 9 labeling (Fig. 2g). Incomplete representation of DNA content could result in partial staining of the corresponding *X. laevis* chromosomal quartet (XI 15+18). An additional reason for lower labeling is the lack of the Xt 9 secondary constriction in *X. laevis* (Fig. 3g). The absence of silver staining on the Xt 9 secondary constriction in *X. tropicalis* could be explained as a

transcriptionally inactive NOR or a cluster of minor rDNA loci.

The remaining painting probes stained chromosomal quartets (Fig. 3a, b, c and d). These and the above mentioned results partially correspond to the reconstruction of a tetraploid karyotype of *X. laevis* based on the similar pattern of high-resolution replication bands (Schmid and Steinlein 1991). The authors grouped all 36 chromosomes in nine quartets (XI 1+2, 3+8, 4+5, 6+9, 7+17, 10+11, 12+13, 14+15 and 16+18). Our results confirmed two of them (XI 1+2 and 6+9), but we were not able to prepare specific painting probes derived from Xt 2 and Xt 4. These probes should stain the remaining four chromosomal pairs (XI 3, 4, 5 and 8). In Tlapakova et al. (2005), FISH-TSA (fluorescence in situ hybridization coupled with tyramide signal amplification) localization of two paralogous *mdh2* genes revealed subcentromeric positions on XI 3 and 8, indicating their common origin. Krylov et al. (2007) performed an analogous experiment with the *mdh2* ortholog in *X. tropicalis*. The cDNA probe labeled a subcentromeric locus on Xt 4. Both results suggest that this chromosome is present in *X. laevis* as the quartet XI 3+8. Therefore, it is probable that the remaining pairs XI 4+5 correspond to Xt 2. To confirm this, it will be necessary to clearly distinguish the morphologically similar chromosomes Xt 2 and 4 using a banding technique applicable to chromosomes dropped on the microdissection membrane.

Laser microdissection of few copies of individual chromosomes coupled with the GenomePlex WGA4/WGA3 amplification and labeling protocol seems to be a powerful tool for the preparation of painting probes. Data from Zoo-FISH experiments could be useful for evolutionary studies in lower vertebrates as well as for the prospective sequencing project of *X. laevis*.

Acknowledgements This work was supported by grant 304/07/P204 from the Grant Agency of the Czech Republic, by grant MSM0021620858 from the Ministry of Education, Youth and Sports of the Czech Republic and by project MZE 0002716202 of the Grant Agency of the Ministry of Agriculture of the Czech Republic.

References

- Chowdhary BP, Raudsepp T (2000) HSA4 and GGA4: remarkable conservation despite 300-Myr divergence. *Genomics* 15(1):102–105
- Chowdhary BP, Raudsepp T (2001) Chromosome painting in farm, pet and wild animal species. *Methods Cell Sci* 23(1–3):37–55
- Courtet M, Flajnik M, Du Pasquier L (2001) Major histocompatibility complex and immunoglobulin loci visualized by in situ hybridization on *Xenopus* chromosomes. *Dev Comp Immunol* 25(2):149–157
- Evans BJ (2008) Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). *Front Biosci* 1(13):4687–4706
- Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC (2004) A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Mol Phylogenet Evol* 33(1):197–213
- Fujiwara A, Abe S, Yamaha E, Yamazaki F, Yoshida MC (1998) Chromosomal localization and heterochromatin association of ribosomal RNA gene loci and silver-stained nucleolar organizer regions in salmonid fishes. *Chromosome Res* 6(6):463–471
- Fukova I, Traut W, Vitkova M, Nguyen P, Kubickova S, Marec F (2007) Probing the W chromosome of the codling moth, *Cydia pomonella*, with sequences from microdissected sex chromatin. *Chromosoma* 116(2):135–145
- Gribble S, Ng BL, Prigmore E, Burford DC, Carter NP (2004) Chromosome paints from single copies of chromosomes. *Chromosome Res* 12(2):143–151
- Griffin DK, Robertson LB, Tempest HG et al (2008) Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution. *BMC Genomics* 14(9):168
- Guttenbach M, Nanda I, Feichtinger W, Masabanda JS, Griffin DK, Schmid M (2003) Comparative chromosome painting of chicken autosomal paints 1–9 in nine different bird species. *Cytogenet Genome Res* 103(1–2):173–184
- Hellsten U, Khokha MK, Grammer TC, Harland RM, Richardson P, Rokhsar DS (2007) Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. *BMC Biol* 25(5):31
- Hotz H, Uzzell T, Berger L (1997) Linkage groups of protein-coding genes in western palearctic water frogs reveal extensive evolutionary conservation. *Genetics* 147(1):255–270
- Khokha MK, Krylov V, Reilly MJ et al (2009) Rapid gynogenetic mapping of *Xenopus tropicalis* mutations to chromosomes. *Dev Dyn* 238(6):1398–1406
- Klein SL, Strausberg RL, Wagner L, Pontius J, Clifton SW, Richardson P (2002) Genetic and genomic tools for *Xenopus* research: the NIH *Xenopus* initiative. *Dev Dyn* 225(4):384–391
- Krylov V, Mácha J, Tlapáková T, Takác M, Jonák J (2003) The c-SRC1 gene visualized by in situ hybridization on *Xenopus laevis* chromosomes. *Cytogenet Genome Res* 103(1–2):169–172
- Krylov V, Tlapakova T, Macha J (2007) Localization of the single copy gene *Mdh2* on *Xenopus tropicalis* chromosomes by FISH-TSA. *Cytogenet Genome Res* 116(1–2):110–112
- Kubickova S, Cernohorska H, Musilova P, Rubes J (2002) The use of laser microdissection for the preparation of chromosome-specific painting probes in farm animals. *Chromosome Res* 10(7):571–577

- Ocalewicz K, Penman DJ, Babiak I (2008) Variation in size and location of the Ag-NOR in the Atlantic halibut (*Hippoglossus hippoglossus*). *Genetica* 133(3):261–267
- Pollet N, Mazabraud A (2006) Insights from *Xenopus* genomes. *Genome Dyn* 2:138–153
- Rab P, Rabova M, Pereira CS, Collares-Pereira MJ, Pelikanova S (2008) Chromosome studies of European cyprinid fishes: interspecific homology of leuciscine cytotoxic marker—the largest subtelocentric chromosome pair as revealed by cross-species painting. *Chromosome Res* 16(6):863–873
- Schmid M, Steinlein C (1991) Chromosome banding in Amphibia. XVI. High-resolution replication banding patterns in *Xenopus laevis*. *Chromosoma* 101(2):123–232
- Schmid M, Steinlein C (2001) Sex chromosomes, sex-linked genes, and sex determination in the vertebrate class amphibia. *EXS* 91:143–176
- Schmid M, Haaf T, Schempp W (1985) Chromosome banding in Amphibia. IX. The polyploid karyotypes of *Odontophrynus americanus* and *Ceratophrys ornata* (Anura, Leptodactylidae). *Chromosoma* 91(3–4):172–184
- Shetty S, Griffin DK, Graves JA (1999) Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Res* 7(4):289–295
- Smith JJ, Voss SR (2006) Gene order data from a model amphibian (*Ambystoma*): new perspectives on vertebrate genome structure and evolution. *BMC Genomics* 29(7):219
- Teruel M, Cabrero J, Montiel EE, Acosta MJ, Sánchez A, Camacho JP (2009) Microdissection and chromosome painting of X and B chromosomes in *Locusta migratoria*. *Chromosome Res* 17(1):11–18
- Tlapakova T, Krylov V, Macha J (2005) Localization, structure and polymorphism of two paralogous *Xenopus laevis* mitochondrial malate dehydrogenase genes. *Chromosome Res* 13(7):699–706
- Tymowska J (1973) Karyotype analysis of *Xenopus tropicalis* Gray, Pipidae. *Cytogenet Cell Genet* 12(5):297–304
- Tymowska J, Kobel HR (1972) Karyotype analysis of *Xenopus muelleri* (Peters) and *Xenopus laevis* (Daudin), Pipidae. *Cytogenetics* 11(4):270–278
- Tymowska J, Fischberg M (1982) A comparison of the karyotype, constitutive heterochromatin, and nucleolar organizer regions of the new tetraploid species *Xenopus epitropicalis* Fischberg and Picard with those of *Xenopus tropicalis* Gray (Anura, Pipidae). *Cytogenet Cell Genet* 34(1–2):149–157
- Vitkova M, Fukova I, Kubickova S, Marec F (2007) Molecular divergence of the W chromosomes in pyralid moths (Lepidoptera). *Chromosome Res* 15(7):917–930

RESEARCH ARTICLE

Open Access

Efficient high-throughput sequencing of a laser microdissected chromosome arm

Eva Seifertova^{1*}, Lyle B Zimmerman², Michael J Gilchrist², Jaroslav Macha¹, Svatava Kubickova³, Halina Cernohorska³, Vojtech Zarsky¹, Nick DL Owens², Abdul K Sesay², Tereza Tlapakova¹ and Vladimir Krylov¹

Abstract

Background: Genomic sequence assemblies are key tools for a broad range of gene function and evolutionary studies. The diploid amphibian *Xenopus tropicalis* plays a pivotal role in these fields due to its combination of experimental flexibility, diploid genome, and early-branching tetrapod taxonomic position, having diverged from the amniote lineage ~360 million years ago. A genome assembly and a genetic linkage map have recently been made available. Unfortunately, large gaps in the linkage map attenuate long-range integrity of the genome assembly.

Results: We laser dissected the short arm of *X. tropicalis* chromosome 7 for next generation sequencing and computational mapping to the reference genome. This arm is of particular interest as it encodes the sex determination locus, but its genetic map contains large gaps which undermine available genome assemblies. Whole genome amplification of 15 laser-microdissected 7p arms followed by next generation sequencing yielded ~35 million reads, over four million of which uniquely mapped to the *X. tropicalis* genome. Our analysis placed more than 200 previously unmapped scaffolds on the analyzed chromosome arm, providing valuable low-resolution physical map information for *de novo* genome assembly.

Conclusion: We present a new approach for improving and validating genetic maps and sequence assemblies. Whole genome amplification of 15 microdissected chromosome arms provided sufficient high-quality material for localizing previously unmapped scaffolds and genes as well as recognizing mislocalized scaffolds.

Keywords: *Xenopus*, *Tropicalis*, Chromosomes, Next generation sequencing, WGA, Genetic map

Background

Recently, complete genomes of many important model organisms have been assembled using either Sanger or next generation sequencing such as Solexa (Illumina), Roche 454, SOLiD etc. [1]. However, repetitive elements in higher eukaryotic genomes interfere with assembly of sequence information alone into unified chromosome-scale scaffolds [2]. This obstacle is usually overcome by construction of physical or meiotic linkage maps to provide long-range contiguity. Physical mapping can be accomplished by a variety of methods including restriction analysis of BAC libraries, radiation hybrid panels, and direct visualization of marker positions on chromosomes using fluorescent *in situ* hybridization (FISH). The latter

approach is quite accurate, but only a few markers can be localized in one run. In meiotic linkage mapping, relationships among polymorphic marker sequences are determined by relative frequency of recombination. However, recombination frequency is highly variable, often decreasing near centromeres and high in hotspots, making it difficult to compare genetic and physical distances. In addition, resolution of linkage analysis depends on the type of markers chosen and their abundance.

The diploid amphibian *Xenopus tropicalis* plays a key role in basic biological research. This model system is particularly valuable for studies of early vertebrate embryonic development [3,4], functional genomics [5,6], cell biology [3,7], and vertebrate genome evolution [8]. Its 1.7×10^9 bp genome was sequenced [9] and a genetic map covering its 10 chromosomes was constructed [10]. Two genome assemblies are in wide use, both available on www.xenbase.org. The version 4.1 assembly (v4.1,

* Correspondence: evaseifertova@natur.cuni.cz

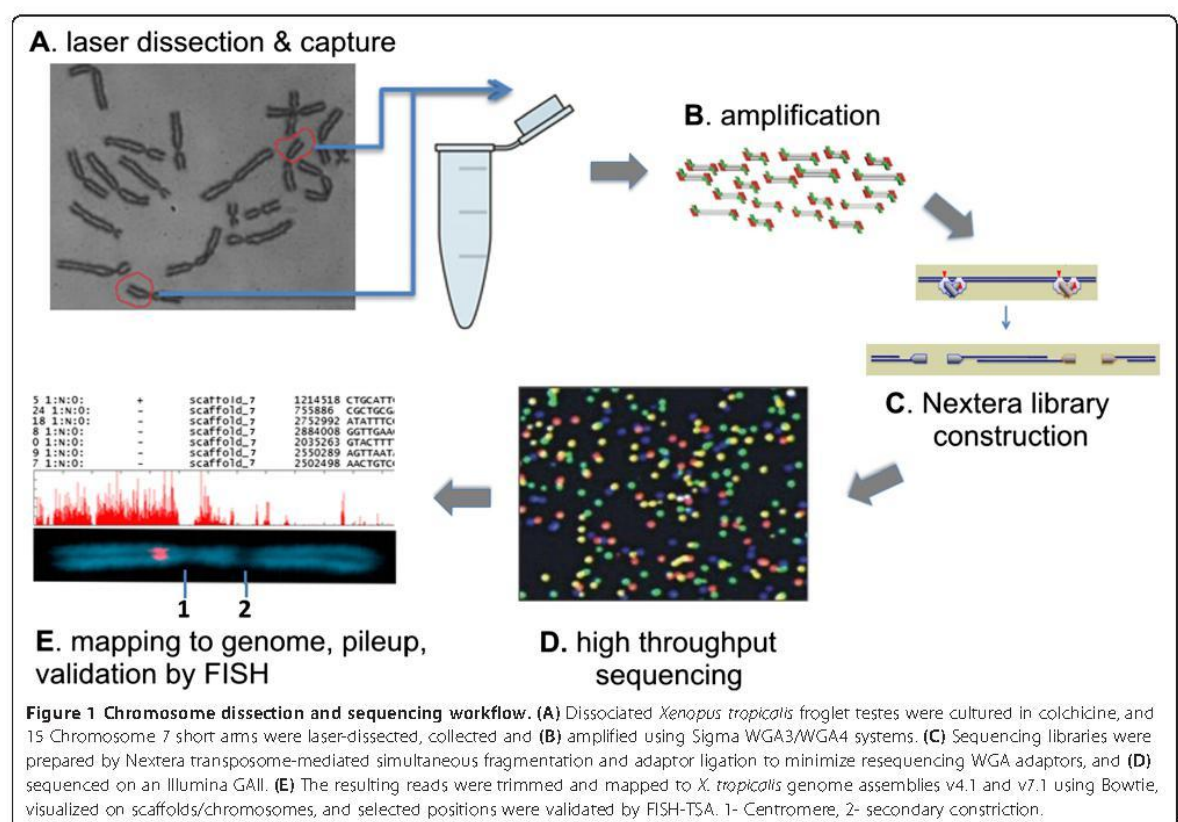
¹Charles University in Prague, Faculty of Science, Vinicna 7, Prague 2 128 44, Czech Republic

Full list of author information is available at the end of the article

Joint Genome Institute) is solely sequence-based and consists of 19,501 scaffolds. A more recent assembly, version 7.1 (v7.1, [9], discussed in [11]), orders reassembled scaffolds using meiotic map and synteny information into a 'main assembly' of 10 chromosome-scale superscaffolds covering ~75% of the genome, with another ~7000 small 'orphan' scaffolds not incorporated into the main assembly. While this long-range assembly is extremely useful, regions assembled by inferring shared gene order with more complete amniote assemblies must be considered provisional, as synteny is not always conserved over large phylogenetic distances. Likewise, the genetic map only locates v4.1 scaffolds covering ~62% of the *X. tropicalis* genome, or about 758 of ~1300 v4.1 [10] scaffolds larger than 100 kb; polymorphic markers were not obtained for the remaining 'unmapped scaffolds'. The largest gaps in the genetic map include the entire short arm of chromosome 2, and a ~15 cM span inside the distalmost marker on the p arm of chromosome 7. Interestingly, the gap on chromosome 7 appears to contain the *X. tropicalis* sex determining locus [12], although an independent marker analysis suggests that there is not a large region of sex-specific sequence [13] which might interfere with meiotic mapping.

To identify sequences within these gaps as well as map and assembly errors, we developed an improved method based on high-throughput sequencing of laser microdissected chromosome arms (Figure 1). Recent technical advances have enabled low cost genome sequencing of nearly any species [14,15]. However, direct sequencing of specific chromosomes or chromosomal regions has only been successful in species where individual chromosomes could be separated by flow sorting (reviewed in [16]). Microdissection of chromosomes has been attempted, but this approach depends on whole-genome amplification due to practical limits on the amount of starting material [17]. In the only published study, sequence resolution was low, probably due to the poor yield and quality of the DNA obtained [18].

Here we used 15 microdissected copies of the short arm of chromosome 7. This small amount of material was then subjected to whole genome amplification (WGA), and sequencing libraries were constructed by transposase-based simultaneous fragmentation and primer insertion, and then sequenced. For WGA, we chose the Sigma GenomePlex single cell kit since it amplified more markers and yielded the highest quantity of DNA relative to other systems [19], and has been successfully



used to amplify microdissected human chromosomes for SNP genotyping [17]. This combination provided excellent read depth for placing previously unmapped scaffolds and genes to the 7p region, as well as recognizing mislocalized scaffolds.

Results

We prepared metaphase chromosomes from primary cell cultures of dissected subadult frog testes [19]. We then laser microdissected 15 copies of the short arm of chromosome 7, which is easily distinguishable due to a secondary constriction in its q arm (Figure 1E). To ensure harvest of the entire p arm, the laser path was targeted between the constriction and centromere (Figure 1E). The extremely small amount of starting material was then amplified by WGA, with a total yield of ~20 µg averaging 500–600 bp in size. For library construction, we wished to minimize resequencing WGA primers added to ends of genomic fragments. To that end, we used a transposase-based simultaneous fragmentation/adaptor ligation method (Nextera, Illumina Inc.) where sequencing primer insertions are biased away from DNA ends. 80 bp reads were then obtained in a single lane of Illumina GAII. Reads were mapped to both versions of *Xenopus tropicalis* assemblies (v4.1 and v7.1) using Bowtie.

Comparison to v4.1 assembly

In total, we obtained 35 million 80 bp reads, 18% of which mapped with a maximum of 2/80 mismatches to at least one location in the v4.1 genome assembly. Of these, 3,900,340 (11% of total) mapped to unique sites in the genome ('hits'; shown in Additional file 1). Despite our efforts to minimize resequencing WGA primers, 30% of non-mapping reads contained at least 11 bases of either Illumina primer or the proprietary Sigma WGA sequence. The remaining non-mapping reads could be either contaminated by shorter stretches of primer or reflected misassembled or missing regions of the genome. Of the uniquely-mapped reads, nearly 70% mapped to scaffolds previously localized to the p arm of chromosome 7 by the genetic map [10], with the majority of those scaffolds showing high read/kb values (see Additional file 2). Since the published genetic map contains large gaps on chromosome 7, chromosome 2 and elsewhere, we also collated available cytogenetic map data from Fluorescence In Situ Hybridization with Tyramide Signal Amplification (FISH-TSA) analysis using cDNA probes [8,10,20], and evaluated read density on scaffolds anchored by known physical location of probes. Read density was congruent with the locations of all 90 physically-mapped scaffolds, including 4 markers in the 7p region (Figure 2, Additional file 3). Near the centromere (genetically mapped to 69 cM +/- 1 cM [12]), this ratio drops from approximately 70 to 17 uniquely-mapped reads/kb, probably due to increased

repeat density diluting unique sequence in centromeric scaffolds. We conservatively used the hit ratio of 17 unique reads/kb observed for Scaffold_298, localized by FISH to just under the centromere [10], as the actual border giving the lowest acceptable hit/kb threshold. All scaffolds with lower than 17 hits/kb were considered false positives (2164 scaffolds containing 6.3% of hits, Table 1). The area between 72 and 73 cM contains scaffolds with around 1 read/kb indicating the laser cutting path. Below this borderline the hit/kb ratio rapidly falls more than 100x.

3.47% of unique hits (135,487) were found in three scaffolds localized by the meiotic linkage map to non-7p regions. Two of these scaffolds were placed on the q arm of chromosome 7 in the linkage map despite relatively high read/kb values in our analysis (scaffold_598, 74.32 cM, 64 hits/kb and scaffold_1153, 96 cM, 57 hits/kb). The third scaffold (scaffold_302) mapped to the q arm of chromosome 8 (38.54 cM, 59 reads/kb). Since the high hit/kb ratios of these scaffolds are similar to those previously mapped to 7p region by FISH-TSA (scaffold_827, 75 reads/kb; scaffold_266, 31 reads/kb; scaffold_271, 79 reads/kb; and scaffold_75, 50 reads/kb) [10], we hypothesize that these sequences should be reassigned to the laser microdissected 7p arm. In total, we identified 231 v4.1 scaffolds not represented on the genetic map (22.5% of unique hits) with a read/kb value higher than 17 (border scaffold_298) which we can assign to the 7p region. 29 of these unmapped scaffolds are larger than 100 kb. On the other hand, 13 scaffolds with markers on 7p in the genetic map bore lower read/kb values than the threshold defined above, more consistent with a non-chromosome 7p location.

Identification of hybrid scaffolds

To identify misassembled (hybrid or broken) scaffolds, uniquely mapped reads were visualized on the *X. tropicalis* 4.1 assembly (Additional file 4). If a whole scaffold were present in the microdissected chromosome part, its full length should be covered by reads. Hybrid scaffolds may show gaps without any chromosome-specific hits, consistent with these sequences deriving from other chromosomal regions. We found that approximately 15% of scaffolds with meiotic map markers on 7p contained gaps in read coverage larger than 100 kb. These gaps in unique read hits are not generally caused by increased repeat density, which is similar in hit-rich and hit-absent areas (Figure 3), and the gaps usually contain genes, also suggesting that they are not due to repetitive sequence. FISH-TSA analysis of three suspected hybrid scaffolds (75, 266 and 270) confirmed assembly discrepancies, with probes from 'hit rich' regions localized to 7p, but cDNA probes from 'hit absent' areas decorating chromosomes 3 and 4 (Figure 3). Meiotic map markers from these

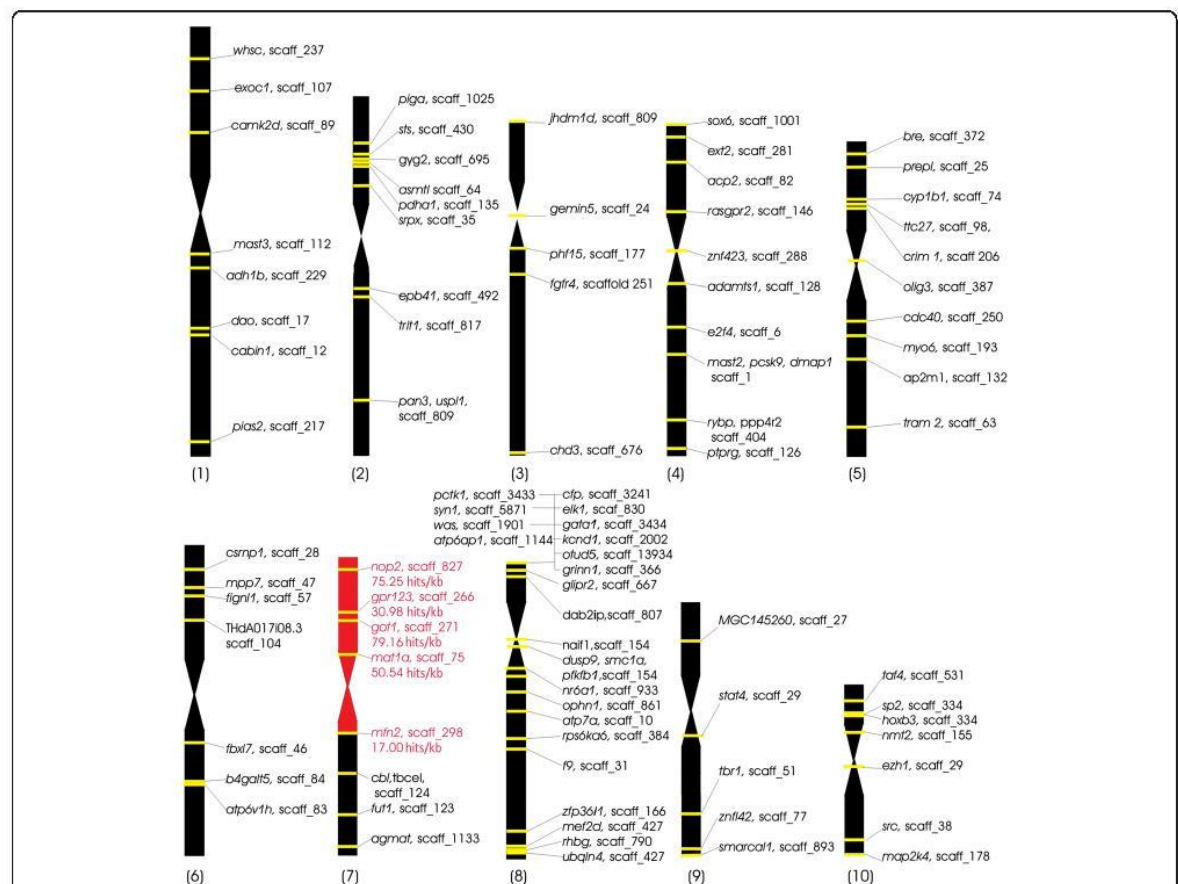


Figure 2 Uniquely-localized read distribution on physically mapped scaffolds. Schematic of *X. tropicalis* karyotype showing genes mapped by FISH-TSA [8,10,20] and cognate v4.1 scaffolds (scaff_number). High read density (17–80 hits/kb) is seen for scaffolds localized to the microdissected region (red region). Genes that physically localized to non-7p regions were all contained by scaffolds with <1 hit/kb (black regions).

scaffolds are also present in 7p and non-7p regions, confirming that these are hybrid 'broken' scaffolds rather than mapping errors. Our analysis thus helps pinpoint bad sequence joins in the scaffold assembly.

Comparison to v7.1 assembly

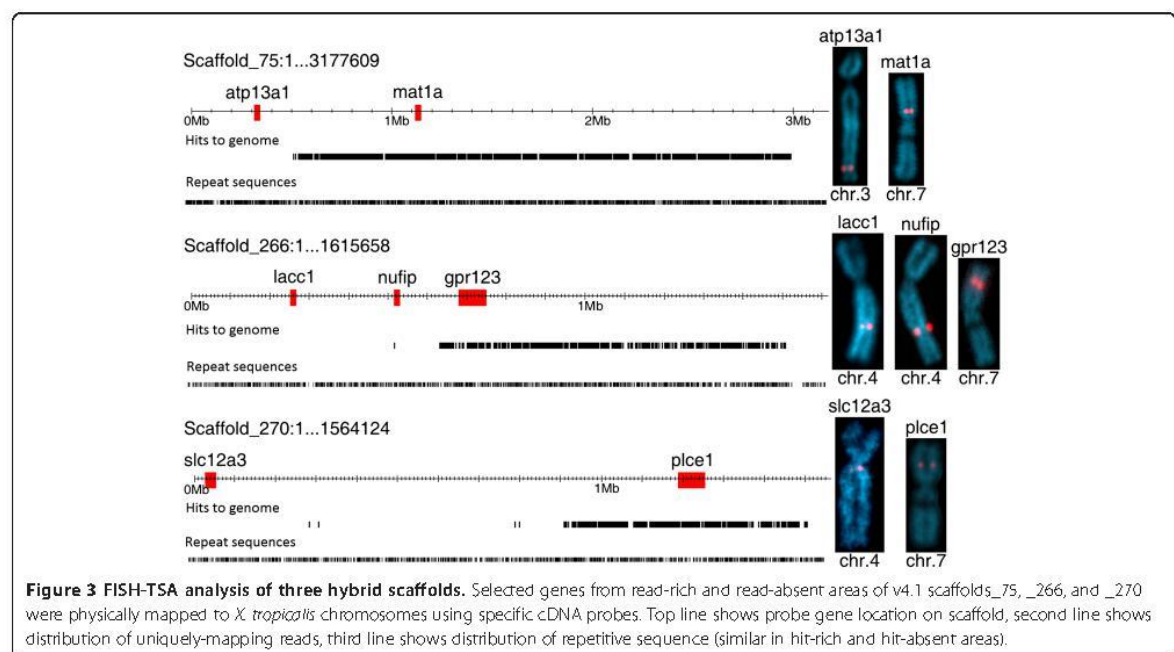
The v7.1 assembly incorporates both sequence information, long-range contiguity from the meiotic map, and gene synteny relationships from amniote genomes. ~75% of the coverage has been ordered provisionally into 10 large superscaffolds corresponding to the 10 *X. tropicalis* chromosomes. In our analysis, 4,489,728 reads placed

uniquely on the v7.1 assembly (see Additional file 5). 80.8% of uniquely-mapping reads localized to scaffold (chromosome) 7, with 91% of these in the 0–60 Mb area roughly corresponding to the 7p region (Figure 4) with an average hit/kb ratio of 29.23. We defined this value (29 reads/kb) as a threshold for hit-positive v7.1 scaffolds. Five large gaps with much lower read density were observed between 0–0.7 Mb, 3.2–4.5 Mb, 5.2–6.7 Mb, 19.7–21.3 Mb and 49.7–55.2 Mb, consistent with areas that were misassembled in v7.1. The 60–65 Mb area is the approximate location of the presumptive laser path, but misassembly in this region is also possible. In the

Table 1 Reads locating uniquely to *Xenopus tropicalis* meiotically-mapped v4.1 scaffolds

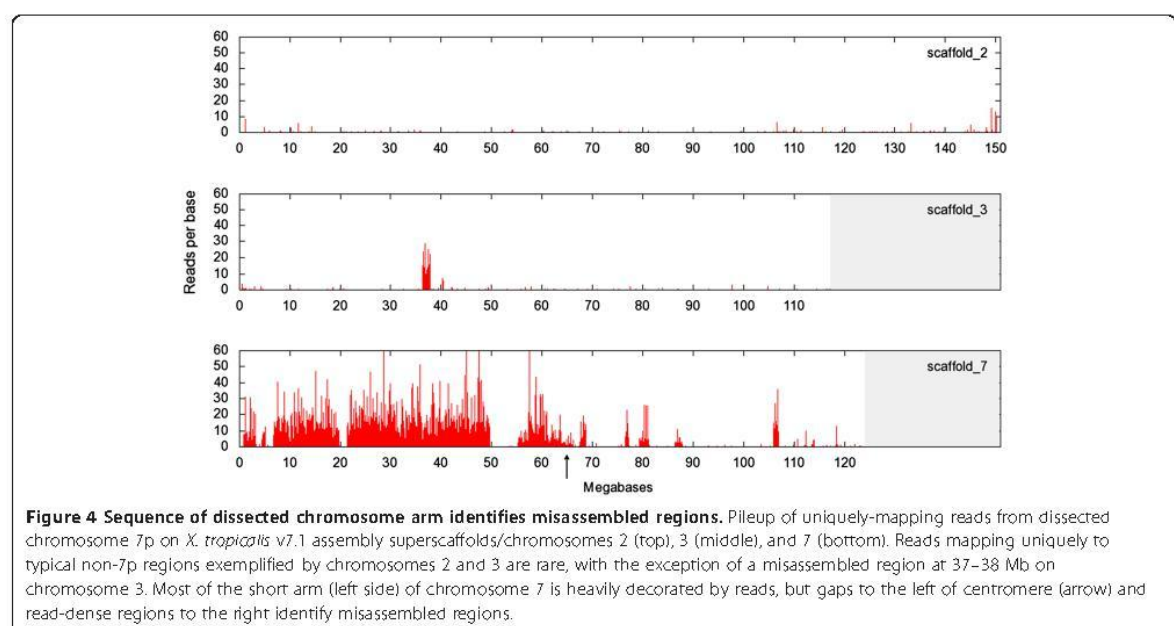
All hits (%)	Hits to 7p v4.1 scaffolds (%)	Hits to above threshold non-7p v4.1 scaffolds (%)	Hits to above threshold unmapped scaffolds (%)	Hits to under-threshold scaffolds (%)
3,900,340(100%)	2,638,303 (67.64%)	135,487 (3.47%)	880,547 (22.58%)	24,6003 (6.3%)

Threshold is set to 17 hit/kb (hit/kb of bordering scaffold_298).



remaining part of scaffold 7 corresponding to 7q (65–120 Mb), we identified four large read-dense regions (67.7–68.8 Mb, 32 hits/kb; 76.5–77.0 Mb, 54 hits/kb; 79.3–81.0 Mb, 17 hits/kb; and 106.0–107.0 Mb, 57 hits/kb) which are likely to be located on the other arm of chromosome 7. Similar sharply-demarcated candidate 7p areas with high read/kb ratios were found also in

chromosome/scaffolds 1 (114.0–115 Mb, 45 hits/kb), 3 (36.7–38.0 Mb, 62 hits/kb, see Figure 4) and 4 (11.0–11.5 Mb, 39 hits/kb, 14.5–15.35 Mb, 34 hits/kb) (shown in Additional file 6). Analysis of ‘orphan’ scaffolds not incorporated into the main chromosomal assembly identified another 14 orphan scaffolds larger than 100 kb showing high hit/kb values comparable to scaffold 7. Of



these, scaffold 35 (27 hit/kb) is covered by high-density reads only on 30% of its length, consistent with a bad sequence join.

Discussion

We present an improved technique for next generation sequencing of laser microdissected chromosome arms. Comparing a previous study [18] using 454 sequencing of DOP-PCR amplified human chromosome arms which yielded similar 80 bp average reads, we obtained approximately 30,000× more reads and 10,000× more unique hits to genome. The increase in usable reads allowed us to set high stringency conditions for mapping (97.5% identity).

As mentioned above, a large portion of the reads that did not map to the genome contained WGA primer sequences, despite precautions to minimize sequencing end fragments by the use of the Nextera system for building the sequencing library. Further optimization of library construction to avoid ends could increase usable read yield considerably.

The yield of mapped reads was comparable to that of a study where wheat flow-sorted chromosome arms were sequenced [21]. The flow sorting approach can collect abundant chromosomal DNA, but can only distinguish a minority of chromosomes in a given karyotype based on size alone [16]. Laser microdissection enables visual control and much greater discrimination among similarly-sized chromosomes, for example sorting by p/q arm ratio in metaphase spreads or using banding techniques. We verified our analysis using available FISH data for 93 relevant v4.1 scaffolds [8,10,20]. All known 7p scaffolds have a high unique read/kb ratio, whereas non-7p scaffolds show a maximum of 1 hit/kb without exceptions (see Figure 2 and Additional file 3). The total fraction of above-threshold reads mapping to chromosome 7p was 93.7% in the v4.1 assembly and 91% in v7.1, comparable to an analysis of human chromosome 19 [22] where 93% of reads mapped to the cognate region.

As mentioned above, the chromosome 7 linkage map contains a large gap between the most distal marker (0 cM) and the next one at 15 cM. We identified 264 scaffolds from the v4.1 assembly in 7p, only 49 of which are represented in the current linkage map. The remaining 215 scaffolds, with a total of size 17 Mb, had either unknown or incorrect positions. Many of these newly

localized scaffolds are likely to be contained in the distal gap as well as smaller gaps elsewhere on the 7p linkage map.

Although v4.1 7p scaffolds show unique read/kb ratios ranging from 17 to 527, these values are sufficient to assign scaffolds to the laser microdissected area. Identification of scaffolds as hybrid or misassembled by virtue of unequal hit coverage was verified by FISH-TSA. Scaffolds close to the laser cut have a lower coverage (17–20 hits/kb), but these hits are evenly distributed across the scaffold. Misassembled scaffolds in the central part of 7p (unaffected by the laser) show sharply uneven distribution of reads, with some areas showing values of ~30–40 hit/kb and others <1 hit/kb.

In our analysis of the v7.1 assembly, 91% of uniquely-mapping reads were to above-threshold (29 hits/kb) scaffolds, a lower percentage than when compared with the v4.1 assembly and map (see Table 2). Our analysis suggests that this is largely due to hits in a few defined regions that were misassembled in v7.1 chromosome-scale superscaffolds 1, 3 and 4, leading to overall below-threshold values for those entire chromosomes/scaffolds. Our analysis also identifies 97 above-threshold orphan scaffolds, contributing an additional 7.2 Mb of sequence to the 7p region.

Analysis of sequence from microdissected chromosome arms identified errors in both available *X. tropicalis* assemblies, with at least 15% of v4.1 scaffolds mapped to 7p scaffolds misassembled. In the v7.1 assembly, we located large regions of 7p sequence which were misassigned to superscaffolds 1, 3, and 4 (Additional file 6). Interestingly, v4.1 scaffolds_75, _266, and _270, identified as hybrids by our analysis, were divided in the v7.1 assembly. However, FISH analysis using probes corresponding to the *atp13a1* and *laccl* genes from hit-absent regions of v4.1 scaffolds_75 and _266 revealed actual locations on chromosomes 3 and 4, respectively, rather than the positions given by the v7.1 assembly on chromosomes/superscaffolds 8 and 2. These results suggest that the v7.1 assembly has successfully identified bad sequence joins, but has not necessarily correctly repaired them in all cases. High throughput sequencing of microdissected chromosomes or chromosomal arms helps to identify such misassembled domains, as well as to assign orphan scaffolds to chromosomal regions. Microdissection and sequencing of particularly problematic areas, such as 7p and 2p, allows sequence domains to be assigned to a specific chromosome arm

Table 2 Reads locating uniquely to *Xenopus tropicalis* v7.1 assembly

All hits (%)	Hits to v7.1 superscaffold 7 (%)	Hits to above-threshold orphan scaffolds (%)	Hits to under-threshold scaffolds (%)
4,489,728	3,627,889	461,438	228,745
(100%)	(80.8%)	(10.26%)	(8.94%)

Threshold set to 29 hit/kb (hit/kb of scaffold 7).

before extrapolating position from meiotic map or synteny data. Improving assembly of the short arm of chromosome 7 is critical for characterizing the sex-determination genes of *Xenopus tropicalis*, which is known to use a different system from the *DM-W* mechanism found in *X. laevis* [23].

Since our method could be combined with chromosome banding, it is likely to be particularly useful for *de novo* assembly of challenging genome projects, such as that of the allotetraploid laboratory model *Xenopus laevis* ($N=18$). In the absence of meiotic or physical map information, correct regional assembly and long-range contiguity would be enhanced by dissecting and sequencing specific chromosomes, all of which in *X. laevis* can be unambiguously distinguished by banding pattern [24].

Conclusions

We have demonstrated feasibility of high-throughput sequencing from as little as 15 microdissected chromosome arms. This approach will be helpful for validating and completing problematic regions in the *X. tropicalis* genome, and can also be used in other species without sequenced genomes for describing gene content in selected chromosomes or providing long-range contiguity. Moreover, the technique is applicable to molecular analysis of isolated chromosomes from small numbers of cells, which is important for investigation of haplotypes or molecular rearrangements in clinical cytogenetics or oncology.

Methods

Chromosomal spreads for laser microdissection

X. tropicalis chromosome nomenclature followed [20]. Metaphase spreads were prepared from euploid primary cell cultures of dissected testes as described in [20] with minor changes. Cells were trypsinized and hypotonized in 38 mM KCl for 5 min. After fixation, cell suspensions were stored overnight at -20°C . For laser microdissection, cells were dropped on a polyethylene naphthalene membrane (P.A.L.M. GmbH, Bernried, Germany) attached to a thin glass slide, allowed to dry, and stained with 3% Giemsa in H_2O for 10 min.

Laser microdissection

Chromosomes were harvested as in [25]. Briefly, 15 copies of the p arm of chromosome 7 were microdissected and collected using a PALM MicroLaser system (P.A.L.M. GmbH, Bernried, Germany) coupled with an inverted microscope (Olympus) under an oil immersion objective (100 \times magnification). Chromosome arms were catapulted by a single laser pulse directly into the cap of a PCR tube containing 4 μL PCR oil. To ensure that the whole short

arm of chromosome 7 was obtained, the laser cut was targeted to the q arm border of the centromere region.

Whole genome amplification

An initial round of whole genome amplification was performed using the WGA4 *GenomePlex Single Cell Kit* (Sigma-Aldrich). Dissected chromosome arms were digested with Proteinase K, followed by library preparation and amplification according to manufacturer's instructions. Primary PCR products were cleaned up using a Qiaquick Gel Extraction Kit (QIAGEN) column. 20 ng of primary WGA4 product was then reamplified using the WGA3 system (*GenomePlex WGA Reamplification Kit*, Sigma-Aldrich) according to our original protocol for preparation of *X. tropicalis* painting probes [26]. The secondary PCR product was purified by ethanol precipitation, yielding approximately 20 μg of DNA fragments averaging ~500–600 bp in size.

High throughput sequencing and library construction

In order to minimize resequencing WGA adaptors at ends of amplified fragments, libraries for high-throughput sequencing were constructed by *in vitro* transposition to simultaneously fragment the DNA and introduce sequencing primer/adaptors using the Nextera DNA sample prep kit (Illumina, Inc.) according to manufacturer's instructions. 50 ng (measured by QuBit, Life Technology) of dissected chromosome amplification product DNA was used with Nextera Illumina-Compatible Enzyme Mix and low molecular weight buffer to generate libraries with fragment size of ~200 to 400 bp (including the 135 bp adapter sequence), and enriched by limited-cycle PCR. Library quality was determined by QuBit, Agilent Bioanalyser and QT-PCR using KAPA library quantification kit (KAPABiosystems, Boston, USA) before loading 6.5 pM on a lane of a GAII flow cell for sequencing 80 bp single reads.

Data analysis

Reads were mapped to *Xenopus tropicalis* v4.1 and v7.1 assemblies (available on www.xenbase.org) [27] using Bowtie (<http://bowtie-bio.sourceforge.net>) [28,29]. Due to prevalence of repeats in the *X. tropicalis* genome, only unique hits with higher than 97.5% identity were selected, using Bowtie parameters $m=1$ and $v=2$. The sequences obtained were counted and analyzed using PERL scripts and Microsoft Excel. Repetitive genomic regions were obtained from the UCSC (browser <http://genome.ucsc.edu/>).

FISH-TSA (Fluorescence In Situ Hybridization with Tyramide Signal Amplification).

Metaphase spreads for FISH-TSA analysis were prepared from the same euploid testes cell cultures as for

the laser microdissection procedure, using the FISH-TSA protocol described in [30].

Additional files

Additional file 1: Reads mapping uniquely to v4.1 assembly. Table showing all unique hits to v4.1 scaffolds, ordered by hit/kb ratio.

Additional file 2: Reads localizing uniquely to meiotically-mapped chromosome 7 scaffolds. Chromosome 7 meiotic map with polymorphic markers and associated *Xenopus tropicalis* v4.1 scaffolds are shown with hits/kb from dissected chromosome 7p sequence. Read density decreases sharply near 72 cM, just beneath centromere at 69 cM +/- 1 cM.

Additional file 3: Uniquely-mapped read distribution on physical map. All genes previously mapped by FISH-TSA [8,10,20] are shown with associated chromosome/linkage group (chrom/LG), meiotic map position, cognate v4.1 scaffolds with hit/kb ratio, probe size (Ampl. Length), relative distance from centromere (RDC), and v7.1 position. All regions with high read density (17–80 hits/kb, marked in red) are located on 7p. FISH-TSA probes in all other regions (black) have <1 hit/kb.

Additional file 4: Reads mapping uniquely to v4.1 scaffolds. The file includes all v4.1 scaffolds larger than 100 kb with hit/kb>17. Every panel contains one scaffold with its length scale, position of all genes, and visualization of reads.

Additional file 5: Reads mapping uniquely to v7.1 assembly. Table showing all unique hits to v7.1 scaffolds. Chromosome-scale superscaffolds are shown top right, with orphan scaffolds on left ordered by decreasing unique reads/kb ratio.

Additional file 6: Read distribution on v7.1 scaffolds. The file includes all scaffolds larger than 100 kb with unique reads/kb>29.23, including scaffold 35 and regions with high coverage in superscaffolds 1, 2, 3, 4 and 8. Scaffolds are ordered by ascending scaffold number. Superscaffold 7 is divided into several smaller parts. Each scaffold shows length scale and reads.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

VK, ES and LZ designed the project; ES, SK, HC, AS, and TT performed the experiments; ES, NO, VZ & MG analyzed the data; ES, VK and LZ wrote the paper. All authors read and approved the final manuscript.

Acknowledgement

We are grateful to Joachim Kurth (NIMR Division of Developmental Biology) for technical assistance. The research was funded by grants from the Grant Agency of Charles University in Prague 407311; Grant Agency of Czech Republic P502/11/P522; Ministry of Education, Youth and Sports of Czech Republic MSM0021620858; UNCE 204013; SVV 265211; Ministry of Agriculture of the Czech Republic MZE 0002716202 and UK Medical Research Council U117560482 and U117597137.

Author details

¹Charles University in Prague, Faculty of Science, Vinicna 7, Prague 2 128 44, Czech Republic. ²MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK. ³Veterinary Research Institute, Hudcova 70, Brno 621 00, Czech Republic.

Received: 12 February 2013 Accepted: 24 May 2013

Published: 28 May 2013

References

- Pagani L, Liolios K, Jansson J, Chen IM, Smirnova T, Nostrat B, Markowitz VM, Kyrpides NC: The Genomes OnLine Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2012, **40**:D571–D579.

- Barthelson R, McFarlin AJ, Rounsley SD, Young S: Plantagora: modeling whole genome sequencing and assembly of plant genomes. *PLoS One* 2011, **6**:e28436.
- Harland RM, Grainger RM: *Xenopus* research: metamorphosed by genetics and genomics. *Trends Genet* 2011, **27**:507–515.
- Kashiwagi K, Kashiwagi A, Kurabayashi A, Hanada H, Nakajima K, Okada M, Takase M, Yaita Y: *Xenopus tropicalis*: an ideal experimental animal in amphibia. *Exp Anim* 2010, **59**:395–405.
- Amaya E: *Xenomics*. *Genome Res* 2005, **15**:1683–1691.
- Carruthers S, Stemple DL: Genetic and genomic prospects for *Xenopus tropicalis* research. *Semin Cell Dev Biol* 2006, **17**:146–153.
- Abu-Daya A, Khokha MK, Zimmerman LB: The hitchhiker's guide to *Xenopus* genetics. *Genesis* 2012, **50**:164–175.
- Macha J, Teichmanová R, Sater AK, Wells DE, Tlapáková T, Zimmerman LB, Krylov V: Deep ancestry of mammalian X chromosome revealed by comparison with the basal tetrapod *Xenopus tropicalis*. *BMC Genomics* 2012, **13**:315.
- Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L, et al: The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 2010, **328**:633–636.
- Wells DE, Gutierrez L, Xu Z, Krylov V, Macha J, Blankenburg KP, Hitchens M, Bellot LJ, Spivey M, Stemple DL, et al: A genetic map of *Xenopus tropicalis*. *Dev Biol* 2011, **354**:1–8.
- Gilchrist MJ: From expression cloning to gene modeling: the development of *Xenopus* gene sequence resources. *Genesis* 2012, **50**:143–154.
- Olmstead AW, Lindberg-Livingston A, Degitz SJ: Genotyping sex in the amphibian, *Xenopus (Silurana) tropicalis*, for endocrine disruptor bioassays. *Aquat Toxicol* 2010, **98**:60–66.
- Bewick AJ, Chain FJ, Zimmerman LB, Sesay A, Gilchrist MJ, Owens ND, Seifertova E, Krylov V, Macha J, Tlapáková T, et al: A large pseudoautosomal region on the sex chromosomes of the frog *Silurana tropicalis*. *Genome Biol Evol* 2013. Epub ahead of print.
- Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, et al: The sequence and de novo assembly of the giant panda genome. *Nature* 2010, **463**:311–317.
- Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg LA, Bouffard P, Butt DW, Crasta O, Crooijmans RP, et al: Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol* 2010, **8**:e1000475.
- Doležel J, Vrána J, Šafář J, Bartoš J, Kubaláková M, Šimková H: Chromosomes in the flow to simplify genome analysis. *Funct Integr Genomics* 2012, **12**:397–416.
- Ma L, Xiao Y, Huang H, Wang Q, Rao W, Feng Y, Zhang K, Song Q: Direct determination of molecular haplotypes by chromosome microdissection. *Nat Methods* 2010, **7**:299–301.
- Weise A, Timmermann B, Giabherr M, Werber M, Heyn P, Kosyakova N, Liehr T, Neitzel H, Konrat K, Bommer C, et al: High-throughput sequencing of microdissected chromosomal regions. *Eur J Hum Genet* 2010, **18**:457–462.
- Höckner M, Erdel M, Spreiz A, Utermann G, Kotzot D: Whole genome amplification from microdissected chromosomes. *Cytogenet Genome Res* 2009, **125**:98–102.
- Khokha MK, Krylov V, Reilly MJ, Gall JG, Bhattacharya D, Cheung CY, Kaufman S, Lam DK, Macha J, Ngo C, et al: Rapid gynogenetic mapping of *Xenopus tropicalis* mutations to chromosomes. *Dev Dyn* 2009, **238**:1398–1346.
- Hernandez P, Martis M, Dorado G, Pfeifer M, Gálvez S, Schaaf S, Jouve N, Šimková H, Valárik M, Doležel J, Mayer KF: Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J* 2012, **69**:377–386.
- Yang H, Chen X, Wong WH: Completely phased genome sequencing through chromosome sorting. *Proc Natl Acad Sci U S A* 2011, **108**:12–17.
- Yoshimoto S, Okada E, Umemoto H, Tamura K, Uno Y, Nishida-Umehara C, Matsuda Y, Takamatsu N, Shiba T, Ito M: A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc Natl Acad Sci U S A* 2008, **105**:2469–2474.
- Schmid M, Steinlein C: Chromosome banding in Amphibia. XVI. High-resolution replication banding patterns in *Xenopus laevis*. *Chromosoma* 1991, **101**:123–132.
- Kubickova S, Cernohorska H, Musilova P, Rubes J: The use of laser microdissection for the preparation of chromosome-specific painting probes in farm animals. *Chromosome Res* 2002, **10**:571–577.

26. Krylov V, Kubickova S, Rubes J, Macha J, Tlapakova T, Seifertova E, Sebkova N: Preparation of *Xenopus tropicalis* whole chromosome painting probes using laser microdissection and reconstruction of *X. laevis* tetraploid karyotype by Zoo-FISH. *Chromosome Res* 2010, **18**:431-439.
27. Xenbase: *Xenopus laevis* and *Xenopus tropicalis* biology and genomics resource. [www.xenbase.org]
28. Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009, **10**:R25.
29. Bowtie: an ultrafast, memory-efficient short read aligner. <http://bowtie-bio.sourceforge.net/>.
30. Krylov V, Tlapakova T, Macha J: Localization of the single copy gene Mdh2 on *Xenopus tropicalis* chromosomes by FISH-TSA. *Cytogenet Genome Res* 2007, **116**:110-112.

doi:10.1186/1471-2164-14-357

Cite this article as: Seifertova et al.: Efficient high-throughput sequencing of a laser microdissected chromosome arm. *BMC Genomics* 2013 **14**:357.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



A Large Pseudoautosomal Region on the Sex Chromosomes of the Frog *Silurana tropicalis*

Adam J. Bewick¹, Frédéric J.J. Chain², Lyle B. Zimmerman³, Abdul Sesay³, Michael J. Gilchrist³, Nick D.L. Owens³, Eva Seifertová⁴, Vladimir Krylov⁴, Jaroslav Macha⁴, Tereza Tlapakova⁴, Svatava Kubickova⁵, Halina Cernohorska⁵, Vojtech Zarsky⁴, and Ben J. Evans^{1,*}

¹Department of Biology, McMaster University, Hamilton, Ontario, Canada

²Max Planck Institute for Evolutionary Biology, Plön, Germany

³MRC National Institute for Medical Research, London, United Kingdom

⁴Charles University in Prague, Faculty of Science, Czech Republic

⁵Veterinary Research Institute, Brno, Czech Republic

*Corresponding author: E-mail: evansb@mcmaster.ca

Accepted: May 3, 2013

Data deposition: This project has been deposited at GenBank under the accession number SRP022004.

Abstract

Sex chromosome divergence has been documented across phylogenetically diverse species, with amphibians typically having cytologically nondiverged (“homomorphic”) sex chromosomes. With an aim of further characterizing sex chromosome divergence of an amphibian, we used “RAD-tags” and Sanger sequencing to examine sex specificity and heterozygosity in the Western clawed frog *Silurana tropicalis* (also known as *Xenopus tropicalis*). Our findings based on approximately 20 million genotype calls and approximately 200 polymerase chain reaction-amplified regions across multiple male and female genomes failed to identify a substantially sized genomic region with genotypic hallmarks of sex chromosome divergence, including in regions known to be tightly linked to the sex-determining region. We also found that expression and molecular evolution of genes linked to the sex-determining region did not differ substantially from genes in other parts of the genome. This suggests that the pseudoautosomal region, where recombination occurs, comprises a large portion of the sex chromosomes of *S. tropicalis*. These results may in part explain why African clawed frogs have such a high incidence of polyploidization, shed light on why amphibians have a high rate of sex chromosome turnover, and raise questions about why homomorphic sex chromosomes are so prevalent in amphibians.

Key words: sex chromosome, pseudoautosomal region, recombination, sex determination, African clawed frogs, *Xenopus tropicalis*.

Introduction

Sex can be advantageous because it decouples beneficial from deleterious mutations via recombination, which increases the variance in fitness effects of linked mutations, and thus the efficiency with which natural selection operates. In species with genetic sex determination, developmental differences between the sexes are initiated by genetic differences between the sex chromosomes. In some lineages, the genes responsible for triggering sex determination vary, and the sex chromosomes (which carry the sex-determining region) are routinely reassigned from one to another ancestral pair of autosomal chromosomes (Fridolfsson et al. 1998; Ross et al. 2009; Evans et al. 2012; Pease and Hahn 2012). Ironically,

suppression of recombination within a sex-specific region is often favored by natural selection, lest a sex-specific, sex-determining allele loses its sex specificity.

The origin of sex chromosomes could be initiated by sexual antagonism (van Doorn and Kirkpatrick 2007), and in many species, this is associated with cessation of recombination between a portion of the sex chromosomes that makes possible unisexual inheritance of a key genomic region that triggers sex determination. Cessation of recombination between the sex chromosomes can be achieved by reducing or eliminating homology (Charlesworth 1991), for example, through point mutations, inversion, deletion, or insertion of DNA. Strikingly, the extent of the nonrecombining region may increase overtime,

© The Author(s) 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Table 1Genomic Regions of *Silurana tropicalis* That Are Putatively Sex Linked Based on Linkage Study of Olmstead et al. (2010) and Sequencing of Chromosome Arm 7p by Seifertova et al. (submitted)

AFLP	Recombination	v4	v7.1	7p?	Portion Sex Linked
E33.M72.143	0	605:241571-241691	7:4966286-4966166	Yes	4435335-5175370
E33.M81.275	0	494:27646-27898	No hits	—	NA
E33.M90.327	0	No hits	211:76840-76535	Yes	ALL
E38.M93.218	0	953:138210-138402	278:109297-109490	No	ALL
No name	0	494:31633-32115	No hits	—	NA
No name	0	494:27541-27902	No hits	—	NA
No name	0	379:817889-818000	2:149826489-149826600	Yes	149787496-150105127
No name	0	736:292586-293067	78:248343-247864	Yes	ALL
No name	0.3	605:245039-245621	7:4963243-4962661	Yes	4435335-5175370
No name	0	605:116800-117215	22:1040592-1040177	Yes	ALL
E40.M52.572	0.4	859:57522-58049	7:3195527-3196069	Yes	76939-3370464
E33.M61.177	0.5	1778:6156-6312; 1778:9971-9815	144:136541-136385	Yes	ALL
E33.M61.177	0.5	1778:6156-6312; 1778:9971-9815	662:22621-22465	No	NONE because 114 is on 7p
No name	0.5	810:261995-262744	94:264236-264985	Yes	ALL
No name	0.5	810:261995-262744	94:263563-263460	Yes	ALL
E32.M94.406	1.9	Multiple hits	22:58856-59241	Yes	ALL
E37.M52.423	1.9	810:325559-325959	144:95531-95931	Yes	ALL
E37.M52.423	1.9	1151:130316-130719	144:95931-95531	Yes	ALL
E41.M83.506	1.9	810:276803-277288	94:279867-279382	Yes	ALL
E32.M35.552	2.6	Multiple hits	No hits	—	NA
E37.M60.232	2.6	6092: 2392-2601	7931:611-828	Yes	ALL
E32.M59.335	2.9	735:292141-292443	7:7903155-7902853	Yes	6687308-9940823

Note.—NA, not applicable. AFLP refers to the name of the AFLP from Olmstead et al. (2010) if provided. Recombination refers to the recombination rate with the sex-determining locus from that study. Scaffold and position of AFLPs are provided for genome assembly version 4.0 (v4) and 7.1 (v7.1). Sex-linked portions that were included in categories in tables 2 and 3 based on the level of recombination (Portion sex linked, with NA meaning not applicable) either refer to base pair positions of a contig within a larger scaffold that is not interrupted by unknown sequence or the entire scaffold was assumed to be sex linked (ALL).

of divergence of the sex-specific region, and to compare molecular evolution and expression of sex-linked and nonsex-linked genes. To this end, we used restriction-site associated DNA (RAD) tags (Baird et al. 2008), a reduced representation next-generation sequencing approach, to genotype millions of homologous nucleotide positions in male and female individuals including positions that are monomorphic in both sexes, polymorphic in one or both sexes, and positions in which a genotype inference (i.e., homozygous or heterozygous) was only possible in one sex due either to sex specificity of the genotyped position or differences in coverage of that position between the sexes. The RAD tag approach produces sequences of thousands of small regions that are adjacent to a rare cutting restriction enzyme site. Because the sequenced portions of the genome are associated with restriction enzyme sites, many homologous sequences are obtained from multiple individuals. Missing data among individuals can arise in unusual cases where mutation generates polymorphism in the presence or absence of the restriction enzyme sites or because of variation among individuals in the depth of sequencing coverage for a particular region. Our analysis incorporated information on sex-linked regions from Olmstead et al. (2010), information from a laser-dissected chromosome arm 7p from a male individual (Seifertova et al., submitted),

which is linked to the sex-linked region identified by Olmstead et al. (2010), and the most recent genome assembly (version 7.1, reference accession PRJNA12348). This study thus provides, for the first time, a comprehensive perspective on the extent of sex chromosome divergence in this species by evaluating the distribution of homozygous and heterozygous genotypes, molecular evolution, and gene expression of sex chromosomes in the context of the rest of the genome.

Materials and Methods

Four female and four male *S. tropicalis* individuals were obtained from Xenopus Express (Brooksville, FL). Sex was confirmed by dissection and species assignment achieved by comparing between 809 and 812 bp of mitochondrial DNA sequence from a portion of the 16S gene from each sample to homologous sequence data from all other known species of African clawed frog (Evans et al. 2011). We performed a phylogenetic analysis on these 8 sequences, 27 sequences from individuals used in the polymerase chain reaction (PCR) screen detailed earlier, all *Silurana* sequences from Evans et al. (2004), 6 *S. tropicalis* samples from Ghana (obtained from tissue archive at the Burke Museum, University of Washington, accession numbers UWBM5957–8, UWBM5961–63, and

UWBM5969), and sequences from 6 individuals from the “golden” strain used by Olmstead et al. (2010) that were provided by Richard Harland. We used an *X. laevis* sequence from South Africa as an outgroup in this analysis, and the total alignment length was 817 bp. Model selection for phylogenetic analysis was accomplished using MrModeltest2 (Nylander 2004). Phylogenetic analysis was performed with MrBayes version 3.1.2 (Huelsenbeck and Ronquist 2001) using the best-fit model based on the Akaike Information Criterion, with two independent Markov chain Monte Carlo (MCMC) runs, each for 2,000,000 generations. Convergence of the MCMC runs on the posterior distribution was assessed by inspecting parameter trends and effective sample sizes using Tracer version 1.5 (Rambaut and Drummond 2007). Based on these analyses, a burn-in of 500,000 generations was discarded before constructing a consensus tree with MrBayes.

Genomic DNA was extracted from liver using QIAGEN DNeasy kit, purified using QIAGEN’s spin purification protocol, and RAD tag library preparation performed by Floragenex, Inc (Eugene, OR). For each individual, two libraries were generated—one used the restriction enzyme *SbfI* and another used *NotI*. The RAD tag libraries were multiplexed on three Illumina flow cells using individual barcodes, and Illumina sequencing was performed at the University of Oregon. These data have been deposited in GenBank (accession number SRP022004).

Illumina sequence reads were sorted by barcode with RADtools v1.2.4 using the “fuzzy_MID” option, which assigns reads with barcode errors to the nearest barcode (Baxter et al. 2011). Data from each individual were independently aligned to the *S. tropicalis* version 7.1 genome using bwa-0.6.2 (Li and Durbin 2009) and samtools.0.1.18 (Li et al. 2009). The “MarkDuplicates” function in picard (<http://picard.sourceforge.net>) was used to mark putative PCR-amplified duplicates, which were then excluded from the genotyping analysis with an aim of minimizing genotyping error. The Genome Analysis Toolkit (GATK) version 2.2-15 was then used to realign indels using the “RealignerTargetCreator” and “IndelRealigner” functions (McKenna et al. 2010; DePristo et al. 2011). The “FixMateInformation” function of picard was then used to adjust mate pair alignments.

Following “Best Practices” guidelines on the GATK website and forum (<http://gatkforums.broadinstitute.org/>) for analysis of genomes that lack known single-nucleotide polymorphisms (SNPs), the “UnifiedGenotyper,” “BaseRecalibrator,” and “PrintReads” functions of GATK were used to iteratively genotype, recalibrate base quality scores, and generate new input (bam) files, using the genotype files generated from “UnifiedGenotyper” as known polymorphic positions to be ignored for base recalibration in each iteration. Convergence was reached by the fifth iteration, in that variable positions recovered from this analysis were 99.8% identical to those from the fourth iteration. The “VariantFiltration” and “SelectVariants” functions of GATK were then used to

identify and exclude genotyped positions that 1) were within 10 bp of an insertion/deletion, 2) had a Phred genotype quality score (Ewing and Green 1998) of less than 30, which means that we removed positions that had a probability of error of greater than 0.001, or 3) had more than one-tenth of the reads mapping equally well to another position and where there were at least four of these reads.

The *S. tropicalis* genome assembly 7.1 consists of 7,730 scaffolds aggregated from 55,234 contigs connected by “N”s within each scaffold. The total number of bases is 1,437,594,934, of which 5% ($n = 71,599,926$) are “N”s. This assembly includes 14 large “super scaffolds” that were assembled using meiotic map, synteny, and cytological data, corresponding to the 10 haploid chromosomes, with some chromosomes being represented by multiple scaffolds (3a and 3b; 5a and 5b; and 8a, 8b, and 8c). The rest of the scaffolds are “orphan scaffolds” whose chromosomal locations are not yet known. We divided the genomic regions into five mutually exclusive groups based on 1) the inferred level of recombination with the sex-determining region by Olmstead et al. (2010), 2) the linkage groups in the genome assembly 7.1 (table 1), and 3) the results of the Illumina sequencing of the dissected petite arm of chromosome 7 (Seifertova et al., submitted). The first of the five groups (“completely sex linked”) included contigs from assembly 7.1 that contain regions that had no recombination (0%) with the sex-determining region in Olmstead et al. (2010). This means that recombination between an AFLP polymorphism and the sex-determining region was not observed in any of 300 individuals assayed by Olmstead et al. (2010). The second group (“partially sex-linked”) included contigs from assembly 7.1 that contain regions that had a recombination rate of more than 0% and less than 3.0% in Olmstead et al. (2010). The third group (“chromosome 7p”) contained sections of scaffolds in assembly 7.1 that are located on chromosome 7p according to Seifertova et al. (submitted), and not in the “completely sex linked” or “partially linked” categories. The fourth group (“non-7p chromosomes”) contained the remaining sections on the chromosome-scale scaffolds in Assembly 7.1, including the portion of scaffold 7 that did not map to chromosome 7p. The fifth group (“other orphans”) contained orphan scaffolds in assembly 7.1 that 1) have not been linked to a chromosome, 2) have no evidence of sex linkage according to Olmstead et al. (2010), and 3) did not map to chromosome arm 7p according to Seifertova et al. (submitted). More specific information on the scaffold or scaffold portions in each of these groups is provided in table 1.

Genome-Wide Distribution of Genotypes in Female and Male *S. tropicalis*

For genomic regions in each of the six categories described earlier, we tabulated genotype patterns for three scenarios (fig. 1) in 500,000-bp windows across the *S. tropicalis*

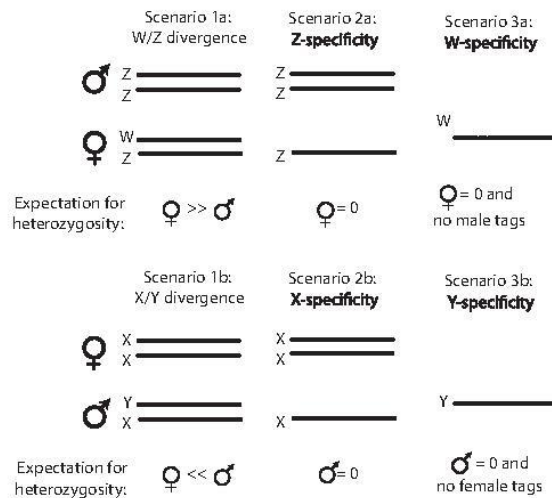


Fig. 1.—Genotypic scenarios for sex-linked regions. Expectations for heterozygosity depend on which sex is heterogametic and the region of the sex chromosome (pseudoautosomal region vs. sex-determining region). Female heterogamy is potentially associated with female-biased heterozygosity (Scenario 1a), male-only heterozygosity (Scenario 2a), or female-only homozygosity with no male genotypes (Scenario 3c). Corresponding scenarios (Scenarios 1b, 2b, and 3b) apply to the opposite sex for male heterogamy.

genome; smaller windows were examined at the ends of scaffolds or when a scaffold was smaller than 500,000 bp. Genotype patterns in each sex (i.e., the distribution of homozygous or heterozygous positions) are relevant to sex chromosome evolution in the following ways. First, divergence between the sex chromosomes due to suppressed recombination generates positions that are either heterozygous in all females and no males (for a ZW sex-determining system) or heterozygous in all males and no females (for an XY sex-determining system). We call these patterns “Scenario 1a” and “Scenario 1b,” respectively (fig. 1). We note that in “Scenario 1a” regions, some positions can also be heterozygous in males due to polymorphism on the Z chromosome, and in “Scenario 1b,” some positions can also be heterozygous in females due to polymorphism on the X chromosome. In any case, in genomic regions consistent with Scenario 1, heterozygosity observed in all samples from one sex is expected to exceed heterozygosity observed in all samples from the other sex. We, therefore, searched for regions with heterozygosity present in all females or in all males. For both of these statistics, we ignored positions that are heterozygous in all genotyped individuals. To account for variation in coverage in males and females, for each window, we divided these counts by the total number of positions in each window for which genotype calls were made in at least one female and at least one male.

Another genotypic scenario for sex chromosomes is that a genomic region may be present only on the Z chromosome (with female heterogamy) or only on the X chromosome (with male heterogamy) (Scenarios 2a and 2b; fig. 1). No counterpart exists on the W chromosome (or Y chromosome) due to deletion, insertion, or divergence. To detect such a genomic region, we searched for regions with heterozygous positions present in one sex but not the other. For such positions, we required a genotype call in at least one individual of each sex but heterozygous calls to be present in only one sex. To account for variation in coverage in males and females, for each window we divided these counts by the total number of positions in each window for which genotype calls were made in at least one female and at least one male.

A third genotypic scenario for sex chromosomes is that a genomic region may be present only on the W chromosome or only on the Y chromosome (Scenarios 3a and 3b, fig. 1). Thus, we searched for positions that had genotype calls only in females (or only in males) and that are all homozygous. To account for variation in coverage, we standardize the counts in each window by the sum of the number of positions in each window for which genotype data are available for 1) at least one female and at least one male, 2) at least one female but no males, and 3) at least one male but no females. Thus, by evaluating these three genotype scenarios in genomic windows across the *S. tropicalis* genome assembly, we attempted to identify genomic windows that either had significantly more heterozygous positions in one sex (Scenario 1), that had heterozygous positions only in one sex (Scenario 2), or that had homozygous positions in only one sex and no homologous genotypes in the other (Scenario 3). Higher values for each ratio are suggestive of genotype patterns characteristic of diverged sex chromosomes.

Expression and Molecular Evolution

As described in Chain et al. (2011), we estimated gene expression levels based on sequences across 26 expressed sequence tag (EST) libraries from the following tissues or developmental stages: egg, gastrula, neurula, embryo, tailbud, tadpole, metamorphosis, adipose tissue, bone, brain, head, heart, intestine, kidney, limb, liver, lung, ovary, oviduct, skeletal muscle, skin, spleen, stomach, tail, testis, and thymus. We summarized patterns of gene expression across EST libraries using the nonindependent “total,” “intensity,” and “evenness” statistics described in Chain et al. (2011). The “total” expression of a gene (T) is the proportion of times that a gene was sequenced in each EST library (L_i) summed across all libraries ($T = \sum L_i$). The “intensity” of expression (I) is the mean expression level from the perspective of a gene and is calculated following this equation: $I = \sum L_i^2 / \sum L_i$. “Evenness” of expression (E) can be thought of as the “effective number” of tissues in which a gene is expressed and is calculated following this equation: $E = T/I$.

For a subset of the sex-linked and nonsex-linked genes, we also calculated the rate ratio of nonsynonymous to synonymous substitutions per site (dN/dS) along the *S. tropicalis* lineage using PAML version 4.5 (Yang 1997). This ratio was calculated using a maximum likelihood model that individually estimates dN/dS for each branch in a phylogeny, following Chain et al. (2011). Our phylogeny was estimated from sequences from *S. tropicalis*, *X. laevis*, and using sequences from another pipid frog (*Pipa carvalhoi* or *Hymenochirus curtipes*) as an outgroup. To avoid undefined values, we added 0.02 to all dS values before calculating dN/dS , following Chain et al. (2011). We made this adjustment a priori by looking only at dS values, to make better use of the data. Because extreme values for dN and dS were occasionally estimated, we excluded from the analysis genes with an estimated dN or dS value above 2, and any genes whose available data comprised less than 100 synonymous positions.

We used a one-sided permutations to test whether the expression and molecular evolutionary statistics differed between genes that either (a) were or (b) were not on the same chromosome as the sex-determining locus. The permutations randomly divided the set of (a + b) values into two groups of size a and b and then calculated the difference between the averages of each group. We repeated this 1,000 times to generate a distribution for the null hypothesis that the values were drawn from the same underlying distribution, and then compared this with the observed differences, which is the test statistic of each test. A significant difference was inferred if the observed difference was greater than 95% of the differences from the permutations. Because these tests are one sided, the operands of the test statistic (i.e., the minuend and subtrahend of each difference) were defined according to specific expectations for sex chromosome degeneration discussed later.

Results

Mitochondrial DNA Variation within *S. tropicalis*, Including the "Golden" Strain

We analyzed phylogenetic relationships among approximately 810-bp region of mitochondrial DNA from the commercially obtained *S. tropicalis* individuals that we used for RAD tags and PCR screens, six individuals from the golden strain used by Olmstead et al. (2010), and several other wild-caught *S. tropicalis* individuals and individuals from other *Silurana* species. An identical mitochondrial DNA sequence was obtained from the six golden strain individuals, one of the samples we used for RAD tag sequencing (a female), 20 of the samples we used for PCR screens (9 females and 11 males), and one individual sampled from Nigeria. Mitochondrial sequences from five samples used in the RAD tag sequencing (2 females and 3 males) were identical to each other and differed from the golden strain sequence by one nucleotide substitution.

Mitochondrial sequences from two other samples used in RAD tag sequencing (1 female and 1 male) and seven samples used in PCR screens (4 females and 3 males) were identical to each other and differed from the golden strain mitochondrial sequence by a different single-nucleotide substitution than the previously mentioned sequence present in five of the RAD tag samples. Mitochondrial sequences from another sample from Nigeria differed from the golden strain mitochondrial sequence by two nucleotide substitutions. Phylogenetic analysis of these and other sequences indicates that the commercially obtained *S. tropicalis* samples used in this study form a well-supported clade that includes two sequences from Nigeria and the six sequences from the golden strain of *S. tropicalis* (fig. 2). This clade is possibly common in individuals east of the Dahomey Gap, a savannah corridor that interrupts the West African rain forest (Salzmann and Hoelzmann 2005).

Reduced Representation Genome-Wide Genotyping from RAD Tags

We used a reduced representation genome sequencing approach called "RAD tags" to sequence many small but homologous portions of the *S. tropicalis* genome in four female and four male individuals. An average of 9,696,525 Illumina reads were mapped in each individual, with the average number of reads mapped per female or per male being 9,445,507 and 9,947,543 reads, respectively. After excluding positions in the reference sequence with no data, within an individual the average depth of coverage was 18.4 reads per position. Genotypes were called for a total of 19,624,824 positions, and 193,199 SNPs (0.98%) were detected. For each position at which at least one genotype was called, an average of 7.14 out of 8 individuals were genotyped.

If *S. tropicalis* has a large female-specific genomic region on the W chromosome, we expected a higher proportion of the Illumina reads from females to map to the genome assembly because this assembly was generated from a female individual. Contrary to this expectation, a slightly higher proportion of reads from males (average per male individual 89.2%, range: 87.8–90.8%) than from females (average per female individual 87.9%; range: 85.0–90.8%) mapped to this genome assembly, arguing against there being a large female-specific region in the *S. tropicalis* genome. Another indication of a large female-specific genomic region on the W chromosome would be a substantially higher number of positions genotyped in females than in males. Out of a total of 19,624,824 positions that were genotyped with high confidence in at least one individual, slightly more genotypes were recovered in females than in males: 912,738 (4.7%) positions were genotyped only in one or more females, and 462,792 (2.4%) positions genotyped only in one or more males. However, in the 10 largest scaffolds, the number of genotype calls in at least one female was consistently 1.1–3.3% higher than the number of genotype calls in at least one male, with

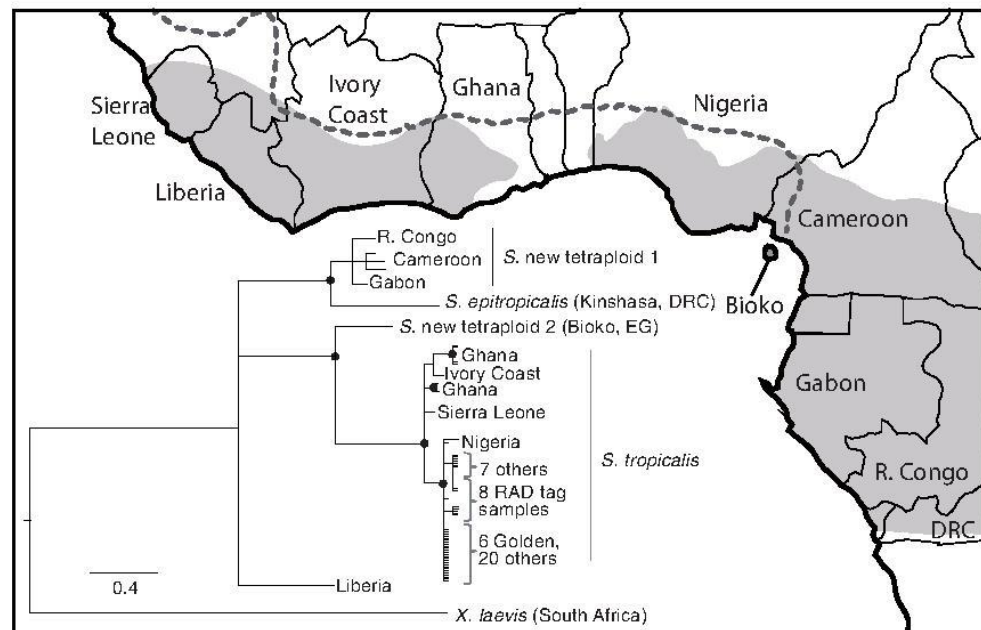


Fig. 2.—Phylogenetic analysis of mitochondrial DNA sequences suggest that the “golden” strain used by Olmstead et al. (2010) and samples used in this study (8 for RAD tag analysis and 27 others for PCR assays) originate from Nigeria. Nodes with $\geq 95\%$ posterior probability are indicated with a black circle. Species names, including those undescribed, follow Evans et al. (2004). Abbreviated country names include the Republic of the Congo (R. Congo), the Democratic Republic of the Congo (DRC), and Equatorial Guinea (EG). The scale bar refers to the number of substitutions per site, gray areas on the map indicate the distribution of tropical forest in West Africa, and the dotted line indicates the approximate distribution of *Silurana tropicalis* inferred by Tinsley et al. (1996).

scaffold 7 having 2.5% more genotype calls in females than males. This suggests that the higher number of unique genotype calls in females is primarily a technical artifact related to differences in coverage among individuals in the RAD tag libraries. The RAD tag data did not provide high-quality genotypes from any positions on 5,721 scaffolds, which together comprise 36,133,437 bp ($\sim 2.1\%$ of the genome).

Genome-Wide Genotype Patterns and Nucleotide Diversity Similar in Males and Females

We searched 500,000-bp windows for various genotypic patterns consistent with sex chromosome divergence expected under female or male heterogamy (fig. 1). In general, this effort failed to identify any regions with a pronounced genotypic signature of sex chromosome divergence expected by female heterogamy (table 2). One exception was a significant excess of windows with female-only homozygous genotypes (Scenario 3a) in orphan scaffolds, but we suspect this was an artifact related to the broader coverage in females. Most notably, portions of linkage groups 2 and 7 that were categorized as “partially sex-linked” and “completely sex-linked” to the sex-determining region based on Olmstead et al. (2010)

did not exhibit a genotypic pattern consistent with degenerate sex chromosomes based on the RAD tag genotypes.

Considerable caution is needed in the interpretation of the average genotype frequencies in genomic windows for the “other orphans” category because in many cases the scaffold is smaller than the window size (500,000 bp), and the resulting truncated genomic windows are therefore expected to have an increased variance in the frequency of various genotypic patterns. Additionally, average genotype frequencies in these genomic windows could fail to detect small scaffolds that have genotypic patterns consistent with sex chromosome divergence. For example, “other orphans” had higher than expected values for Scenarios 2b and 3b, which are consistent with male heterogamy, and this is probably related to the small size of these scaffolds and consequent increase in the sex-specific genotypes in truncated windows for these scaffolds.

Additional insights are gained by examining nucleotide diversity in each sex within 500,000-bp windows. If a portion of the sex chromosomes is substantially diverged, we expected much higher average nucleotide diversity per site in one sex (females for female heterogamy) in genomic windows spanning this diverged region. However, average nucleotide

Table 2
Putatively Sex-Linked Regions Do Not Exhibit Genotypic Hallmarks Expected for Diverged Sex Chromosomes

Region	Number of Genotype Calls	Scenario 1a: All Females Heterozygous	Scenario 1b: All Males Heterozygous	Scenario 2a: All Females Homozygous	Scenario 2b: All Males Homozygous	Scenario 3a: All Females Homozygous, Males No Genotype	Scenario 3b: All Males Homozygous, Females No Genotype
"Non-7p chromosomes"	17,026,460	0.00013 (0.00000–0.00087)	0.00018 (0.00000–0.00123)	0.00124 (0.00000–0.00399)	0.00179 (0.00000–0.00504)	0.04980 (0.00390–0.15189)	0.02670 (0.00050–0.09960)
"Chromosome 7p"	830,811	0.00019	0.00019	0.00072	0.00456	0.10066	0.03760
"Other orphans"	1,618,864	0.00034	0.00040	0.00305	0.00524*	0.33587*	0.18522*
"Partially sex-linked"	132,177	0.00024	0.00014	0.00182	0.00263	0.06098	0.03920
"Completely sex-linked"	16,512	0.00026	0.00003	0.00228	0.00308	0.02848	0.01176

Non—Analysis of 500,000-bp genomic windows indicates that genotype patterns of genomic regions linked to sex do not resemble scenarios expected for sex chromosomes substantially more than other parts of the genome. Numbers indicate average values for three genotypic scenarios depicted in figure 1 and described in Materials and Methods section. For the "Non-7p chromosomes," 95% confidence intervals are in parentheses. Asterisks indicate values that are higher than the 95% confidence intervals from the "non-7p chromosomes."

diversity per site is essentially identical in males and females throughout these scaffolds, including chromosome arm 7p, which is linked to the sex-determining region (fig. 3). To explore the possibility that there could be variation within the RAD tag samples in sex chromosome divergence that corresponds with the three mitochondrial DNA haplotype groups detailed earlier, we explored nucleotide diversity in male and female individuals from each group. This analysis also did not identify a pronounced signature of sex chromosome divergence (supplementary figs. S1 and S2, [Supplementary Material online](#)).

Genotype Patterns Based on Sanger Sequencing

We amplified 65 portions (amplicons) of genomic regions identified by Olmstead et al. (2010) to be linked to the sex-determining region, including 18 and 46 amplicons from "completely sex-linked" and "partially sex-linked" regions, respectively (table 1, [supplementary table S1](#) and [fig. S3](#), [Supplementary Material online](#)). None had female-specific amplifications, allowing us to dismiss Scenario 3 (fig. 1) for all these regions. We sequenced 45 of these amplifications in multiple male and female individuals. SNPs or insertion/deletion polymorphisms were shared between males and females in at least one amplicon for essentially all scaffolds (no polymorphism was observed in amplicons from scaffold 144). This suggests that Scenario 1 is unlikely for these regions, with the caveat being that a heterozygous position could arise in both sexes in a region consistent with Scenario 1 through convergent evolution on the W and Z.

We also used PCR to examine an additional 173 regions that exhibited signs of sex linkage based on our analyses of the RAD tag data, including regions of chromosome 7p and elsewhere as detailed in [supplementary figure S3](#) and [table S1](#), [Supplementary Material online](#). None had sex-specific amplifications, allowing us to dismiss Scenario 3 for all these regions. We sequenced 94 of these amplifications from male and female individuals. Thirty of these were not polymorphic in any of the individuals we sequenced. Forty-eight had polymorphisms shared between males and females, allowing us to conclude that Scenario 1 is unlikely for these regions. Five had polymorphisms in both sexes with none being shared across sexes. Ten had polymorphisms only in females and two had polymorphisms only in males.

Two amplifications were of particular interest. An amplification on scaffold 7 that spanned positions 10,128,301–10,129,920 was highly polymorphic in females but not males, although no polymorphism was fixed in females (out of seven females and three males sequenced; [supplementary table S1](#), [Supplementary Material online](#)). This region failed to amplify in four females and three males. Another amplification, which targeted a region on scaffold 163 had 31 polymorphisms in three females but only one polymorphism in

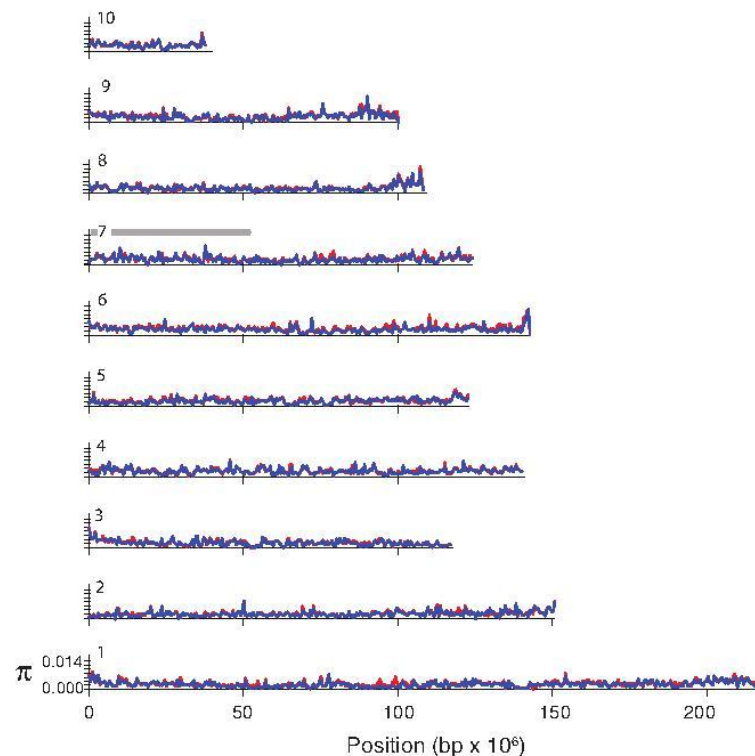


Fig. 3.—Average nucleotide diversity per site (π) in 500,000-bp windows is similar in males (blue) and females (red) throughout much of the *Silurana tropicalis* genome. Plots are labeled with numbers referring to scaffolds 1–10, which collectively comprise approximately 75% of the genome. A gray bar on scaffold/chromosome 7 indicates the petite arm based on the linkage map of Wells et al. (2011), which carries the sex-determining region in the *S. tropicalis* golden strain (Olmstead et al. 2010).

three males, but essentially all the female polymorphisms were present in only one individual.

Gene Expression and Molecular Evolution

Expression was detected in a total of 37,790 transcripts in at least one of the 26 EST libraries we surveyed (table 3). On the basis of studies of recently diverged neosex chromosomes in fruit flies (*Drosophila*) (reviewed in Bachtrog 2013), we expected expression of genes situated near the sex-determining locus to be expressed 1) at a lower total level, 2) higher intensity, 3) lower evenness, and to have 4) higher dN/dS compared with genes in other parts of the genome (i.e., the “non 7p chromosomes”). For the most part, these expectations were not met for genes that were demonstrably very close to the sex-determining region, with the one exception that the intensity of “completely sex-linked” genes were individually significantly higher than the “non 7p chromosomes” ($P < 0.05$, table 3). Evenness of “other orphans” was also significantly lower than “non 7p chromosomes” as was total expression of “other orphans.” dN/dS was significantly higher only in

“chromosome 7p” compared with “non-7p chromosomes” but the magnitude of this difference was small. No expression data were recovered from genes on Scaffold 22, which is tightly linked to the sex-determining region (Olmstead et al. 2010), even though it was 1,156,260-bp long. We examined this scaffold using Xenbase (Bowes et al. 2009) and found that it contained a cluster of olfactory receptors, which (not surprisingly) were not highly expressed in any of the EST libraries we examined.

Discussion

To explore sex chromosome divergence in an amphibian, we used genotype calls from approximately 20 million positions, information about sex linkage, EST databases, and molecular evolutionary analyses to further characterize the sex chromosomes of the Western tropical frog *S. tropicalis*. Phylogenetic analysis of mitochondrial DNA sequences suggests our samples originated in Nigeria, which is also the source of the female individual from which the genome sequence was generated (Hellsten et al. 2010). Additionally, our analysis also

Table 3Average Expression and Molecular Evolution Statistics for *Silurana tropicalis* Genes in Five Genomic Categories

Region	Number of Genes (Expression)	Total	Intensity	Evenness	Number of Genes (dN/dS)	dN/dS
"Non-7p chromosomes"	35,135	0.00065	0.00015	3.19655	9183	0.2702
"Chromosome 7p"	2,114	0.00079	0.00019	3.17054	546	0.2850*
"Other orphans"	260	0.00040*	0.00012	2.59443*	55	0.2670
"Partially sex-linked"	246	0.00090	0.00018	3.37658	71	0.2751
"Completely sex-linked"	35	0.00101	0.00049*	2.56020	8	0.2615

Note.—See Materials and Methods for description of statistics.

*Values that are individually significantly different from the "Non-7p chromosomes" ($P < 0.05$, one-sided permutation tests).

suggests that the golden strain analyzed by Olmstead et al. (2010) is from Nigeria.

Known sequences in the *S. tropicalis* genome sequence assembly version 7.1 comprise approximately 80.4% of the approximately 1.7 Gbp genome, and scaffolds, including "N"s, comprise approximately 84.5% of the genome. Thus, the RAD tag data could not be compared with 15–20% of the genome because of gaps in the genome sequence. Because of variation in coverage, high confidence genotype calls were not made on scaffolds that together comprise an additional 2.1% of the genome. Thus, in this study, we lack information from a nontrivial portion of this genome.

Mindful of these substantial gaps in genome sequence and the uncertainty in linkage relationships among many unassembled (orphan) scaffolds, we leveraged information from a targeted sequencing effort of chromosome arm 7p and also the linkage analysis by Olmstead et al. (2010) to guide our analysis. The dearth of genotypic patterns consistent with divergent sex chromosomes, and particularly patterns that are consistent with female heterogamy (table 2), and the similar level of pairwise nucleotide diversity in males and females throughout the petite arm of chromosome 7 (fig. 3) argues strongly against there being a large sex-specific region of the *S. tropicalis* chromosomes. This inference is consistent with the findings of Uno et al. (2008) who detected no sex differences in C-banded heterochromatin in *S. tropicalis*.

On the basis of studies of fruit flies (reviewed in Bachtrog 2013), we expected genes linked to the sex-determining locus to potentially exhibit lower total expression and higher specificity (i.e., higher intensity and lower evenness as defined in Materials and Methods section). We also expected molecular evolution of these genes to be consistent with relaxed purifying selection. However, on the basis of a small sample size, we only observed a significant increased expression intensity of "completely linked" genes compared with the rest of the genome, with none of these expectations met in "partially sex-linked" genes (table 3). Some of these expectations were also met in orphan scaffolds, which have undetermined linkage relationships with respect to the sex-determining locus, and regions of chromosome arm 7p. It is not clear

that these latter observations are related in any way to linkage to the sex-determining region.

Caveats exist in our interpretation of these data. First, non-recombining portions of the genome tend to accumulate repetitive sequences that can be difficult to sequence and map. For this reason, the sex-specific portion of the *S. tropicalis* genome may be under-represented in the current genome assembly and/or our mapped Illumina reads. Second, it is conceivable that there is polymorphism in the sex-determining mechanism (Olmstead A, personal communication). Polymorphism in genetic sex determination could occur at a single locus wherein multiple, differently functioned sex-determining alleles are segregating at a single locus, which have distinct and not necessarily transitive dominance relationships. Polymorphism in genetic sex determination could also occur at multiple loci distributed on the same or different chromosomes. Sex determination in zebrafish, for example, appears to be orchestrated by genes on different chromosomes (Anderson et al. 2012). Genotypic patterns expected with these types of polymorphisms are unclear and could include a dearth or absence of pronounced sex chromosome divergence. A third caveat to our conclusions is that polymorphism among females could also potentially exist in the extent of divergence between the W and Z chromosomes. Under this scenario, it is conceivable that there could be variation among populations in the extent of recombination along the sex chromosomes and consequently the extent and magnitude of divergence between the sex chromosomes. Further exploration of these possibilities will be assisted by the identification of the sex-determining locus in *S. tropicalis*, the completion of high-quality sequencing and assembly of sex-linked regions, and the exploration of variation within and among populations in sex determination and sex chromosome evolution.

Polyploidization, Dosage Compensation, and Sex Chromosome Turnover

Within a species, the propensity to undergo genome duplication and sex chromosome evolution is potentially interrelated. For example, polyploidization might be less common in species with divergent sex chromosomes where one has

degenerated because, after duplication, the degenerate ancestral sex chromosome would segregate as a new autosomal chromosome, and the resulting homozygous null genotypes could be detrimental (Evans et al. 2012). Sex chromosome degeneration also creates imbalances in allelic copy number between the sexes, which can lead to the evolution of dosage compensation—a factor that is also potentially relevant to polyploid speciation (Orr 1990). Dosage compensation is a process that equalizes expression levels in each sex of a gene that has a different number of alleles in each sex. This could evolve in a species with female heterogamy, for example, through inactivation of one of the Z alleles in males or through upregulation of the Z allele in females. Orr (1990) proposed that dosage compensation in species with a degenerate sex chromosome could act as a barrier to polyploid speciation because dosage compensation would be disrupted when a newly formed triploid individual backcrosses with a diploid parental individual. Our analyses suggest that the sex-specific region of *S. tropicalis* is small, that sex chromosome divergence is minimal, and therefore that dosage compensation associated with degeneration of the sex-specific sex chromosome would have evolved in very few genes or none at all. Together these features of the sex chromosomes may have facilitated (or at least not impeded) polyploidization in *Silurana*, which occurred at least once (reviewed in Evans 2008). Interestingly, the sister genus *Xenopus* has a newly evolved sex-determining gene called *DM-W* (Yoshimoto et al. 2008; Bewick et al. 2011). Species in this group also probably have minimally diverged sex chromosomes and have undergone polyploid speciation multiple times (Evans 2008). Clearly, however, this is not the only consideration in the ability of species to tolerate polyploidization because many amphibian groups that have homomorphic sex chromosomes lack polyploid species.

The extent of sex chromosome degeneration is also relevant to the chances a species experiences future sex chromosome turnover—a change in which pair of chromosomes carries the trigger for sex determination (Charlesworth and Mank 2010). If sex chromosome turnover occurs in a species with a diverged and degenerate sex chromosome, the ancestral degenerate chromosome could segregate autosomally, and some individuals could inherit two copies and be homozygous for degenerate alleles (Charlesworth and Mank 2010). Thus, sex chromosome turnover may be more likely in species that have sex chromosomes that are not substantially degenerated.

If sex chromosome turnover were common, this could maintain homomorphy of sex chromosomes. Recent work on sex chromosomes in African clawed frogs has established nonhomology between the sex chromosomes of *X. laevis* and *S. tropicalis* (Uno et al., submitted) and the recent appearance of a novel sex-determining locus in *X. laevis* (Yoshimoto et al. 2008; Bewick et al. 2011). Thus, it appears that the origin of a new sex-determining gene in *X. laevis* was associated with a

reassignment of sex chromosomes without necessarily involving a change in heterogamy, and that recent sex chromosome turnover can account for sex chromosome homomorphy in this species (Tymowska 1991). In other species, including *S. tropicalis*, it is also possible that the sex-determining mechanism of nondiverged sex chromosomes could be old, but that divergence is prevented by periodic recombination, which possibly could be facilitated by breeding individuals that are phenotypically sex reversed (the “fountain of youth” hypothesis; Perrin 2009). Because we do not yet know the sex-determining gene(s) of *S. tropicalis* or other frogs that might have inherited this sex determination system from a recent common ancestor (e.g., genera *Hymenochirus*, *Pseudhymenochirus*, or *Pipa*), we cannot determine at this time whether turnover or recombination best accounts for the apparent homomorphy of the sex chromosomes of this species. Additional identification of sex-determining genes in amphibians, and analysis of their evolutionary histories and genomic context, is thus an exciting direction for future research.

Supplementary Material

Supplementary figures S1–S3 and table S1 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

The authors thank Richard Harland, Mustafa Khokha, James Evans, and Maura Lane for assistance with samples and information on the “golden” strain of *S. tropicalis*. They thank three anonymous reviewers for helpful comments on an earlier draft of this manuscript. They additionally thank Brian Golding and Sharcnet (www.sharcnet.ca) for access to computational facilities. This work was supported by grants from the National Science and Engineering Research Council of Canada, an Ontario Internal Prestige Scholarship, an Early Researcher Award from the Ontario Ministry of Economic Development and Innovation, and McMaster University. Support for sequencing of the laser dissected of chromosome arm 7p was provided by UK Medical Research Council U117560482 and U117597137, the Grant Agency of Charles University in Prague (407311), the Grant Agency of Czech Republic (P502/11/P522), the Ministry of Education, Youth and Sports of Czech Republic (MSM0021620858), UNCE (204013), SVV (265211), and the Ministry of Agriculture of the Czech Republic (MZE 0002716202).

Literature Cited

- Anderson JL, et al. 2012. Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS One* 7:e40701.
- Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet*. 14:113–124.

- Baird NS, et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376.
- Baxter SW, et al. 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One* 6:e19315.
- Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol.* 24:94–102.
- Bewick AJ, Anderson DW, Evans BJ. 2011. Evolution of the closely related, sex-related genes DM-W and DMRT1 in African clawed frogs (*Xenopus*). *Evolution* 65:698–712.
- Bowes JB, et al. 2009. Xenbase: a *Xenopus* biology and genomics resource. *Nucleic Acids Res.* 36:D761–D772.
- Chain FJ, Dushoff J, Evans BJ. 2011. The odds of duplicate gene persistence after polyploidization. *BMC Genomics* 12:599.
- Charlesworth B. 1991. The evolution of sex chromosomes. *Science* 251:1030–1033.
- Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philos Trans R Soc Lond B Biol Sci.* 355:1563–1572.
- Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95:118–128.
- Charlesworth D, Mank JE. 2010. The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. *Genetics* 186:9–31.
- DePristo MA, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature* 43:491–498.
- Evans BJ. 2008. Genome evolution and speciation genetics of allopolyploid clawed frogs (*Xenopus* and *Silurana*). *Front Biosci.* 13:4687–4706.
- Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC. 2004. A mitochondrial DNA phylogeny of clawed frogs: phylogeography on sub-Saharan Africa and implications for polyploid evolution. *Mol Phylogenet Evol.* 33:197–213.
- Evans BJ, et al. 2011. Description of a new octoploid frog species (Anura: Pipidae: *Xenopus*) from the Democratic Republic of the Congo, with a discussion of the biogeography of African clawed frogs in the Albertine Rift. *J Zool* 283:276–290.
- Evans BJ, Pyron RA, Wiens JJ. 2012. Polyploidization and sex chromosome evolution in amphibians. In: Soltis PS, Soltis DE, editors. *Polyploidy and genome evolution*. Springer Verlag, p. 385–410.
- Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8:186–194.
- Fridolfsson A, et al. 1998. Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proc Natl Acad Sci U S A.* 95:8147–8152.
- Hellsten U, et al. 2010. The genome of the western clawed frog *Xenopus tropicalis*. *Science* 328:633–636.
- Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
- Kamiya T, et al. 2012. A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (*Fugu*). *PLoS Genet.* 8:e1002798.
- Kondo M, et al. 2006. Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome Res.* 16:815–826.
- Kuroiwa A, Ishiguchi Y, Yamada F, Shintaro A, Matsuda Y. 2010. The process of a Y-loss event in an XO/XO mammal, the Ryukyu spiny rat. *Chromosoma* 119:519–526.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li H, et al. 2009. The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics* 25:2078–2079.
- Matsubara K, et al. 2006. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *Proc Natl Acad Sci U S A.* 103:18190–18195.
- McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.
- Nylander JAA. 2004. MrModeltest v2. Uppsala University: Evolutionary Biology Centre.
- Olmstead AW, Lindberg-Livingston A, Degitz SJ. 2010. Genotyping sex in the amphibian, *Xenopus (Silurana) tropicalis*, for endocrine disruptor bioassays. *Aquat Toxicol.* 98:60–66.
- Orr HA. 1990. “Why polyploidy is rarer in animals than in plants” revisited. *Am Nat.* 136:759–770.
- Pease JB, Hahn MW. 2012. Sex chromosomes evolved from independent ancestral linkage groups in winged insects. *Mol Biol Evol.* 29:1645–1653.
- Perrin N. 2009. Sex reversal: a fountain of youth for sex chromosomes? *Evolution* 63:3043–3049.
- Rambaut A, Drummond AJ. 2007. Tracer v1.5. Available from: <http://beast.bio.ed.ac.uk/Tracer>.
- Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. 2009. Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PLoS Genet.* 5:e1000391.
- Salzmänn U, Hoelzmann P. 2005. The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *Holocene* 15:190–199.
- Schmid M, et al. 2010. The chromosomes of Terraranan frogs. *Cytogenet Genome Res.* 130–131:1–568.
- Skaletsky H, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825–837.
- Stöck M, et al. 2011. Ever-young sex chromosomes in European tree frogs. *PLoS Biol.* 9:e1001062.
- Tinsley RC, Loumont C, Kobel HR. 1996. Geographical distribution and ecology. In: Tinsley RC, Kobel HR, editors. *The biology of Xenopus*. Oxford: Clarendon Press, p. 35–59.
- Tsuda Y, Nishida-Umehara C, Ishijima J, Yamada K, Matsuda Y. 2007. Comparison of the Z and W sex chromosomal architectures in elegant crested tinamou (*Eudromia elegans*) and ostrich (*Struthio camelus*) and the process of sex chromosome differentiation in palaeognathous birds. *Chromosoma* 116:159–173.
- Tymowska J. 1991. Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. In: Green DS, Sessions SK, editors. *Amphibian cytogenetics and evolution*. San Diego (CA): Academic Press, p. 259–297.
- Uno Y, et al. 2008. Diversity in the origins of sex chromosomes in anurans inferred from comparative mapping of sexual differentiation genes for three species of the Ranidae and Xenopodinae. *Chromosome Res.* 16:999–1011.
- van Doorn GS, Kirkpatrick M. 2007. Turnover of sex chromosomes induced by sexual conflict. *Nature* 449:909–912.
- Wells DE, et al. 2011. A genetic map of *Xenopus tropicalis*. *Dev Biol.* 354:1–8.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 13:555–556.
- Yoshimoto S, et al. 2008. A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc Natl Acad Sci U S A.* 105:2469–2474.

Associate editor: Judith Mank

4. Shrnutí výsledků a diskuze

4.1 Genetická mapa

Nově vytvořená genetická mapa *X. tropicalis* má dvě části- vazebnou mapu a fyzickou mapu tvořenou FISH markery vytvořenou v naší laboratoři. Vazebná mapa vycházela z assembly 4.1 a s každým mapovaným markerem byla do vazebné mapy přidána délka celého scaffoldu, ve kterém se nacházel.

Vazebná mapa byla vytvořena z celkem 2886 SSLP (Simple Sequence Length Polymorphism) markerů. Původně byly tyto markery vyhledávány ve starší verzi assembly v.2.0, nicméně do následné analýzy byly zahrnuty pouze takové, které měly unikátní sekvenci v assembly v.4.1. Polymorfnní markery byly hledány mezi jedinci kmenů Ivory coast a Nigerian Strains. Experimentálním zpětným křížením bylo získáno 190 jedinců, u kterých byl zjišťován výskyt jednotlivých polymorfií. Vazebná analýza provedená programem JoinMap3 vedla k vytvoření celkem 10 velkých a 4 menších vazebných skupin. Každá vazebná skupina obsahovala 103-506 markerů a jejich délka byla 77-138 centimorganů. 4 menší vazebné skupiny celkem zahrnovaly 54 markerů (6 až 29 na skupinu) a byly 6-40 cM dlouhé. Celková délka vazebné mapy představovala 1345,9 cM, zahrnovala 758 z největších cca 1300 scaffoldů s velikostí přes 100.000 bp a pokrývala minimálně 62% genomu *Xenopus tropicalis* (přibližně 1,055Gb).

Průměrná vzdálenost mezi markery činila u jednotlivých vazebných skupin 0,27-0,75. 1cM odpovídal průměrně vzdálenosti 581 kb, nicméně tato hodnota velmi kolísala (od 4kb/cM do 3000kb/cM). Především v oblasti centromer byla tato hodnota velmi vysoká, nicméně i některé scaffoldy nacházející se ve velké vzdálenosti od centromery měly tuto hodnotu zvýšenou.

Přiřazení jednotlivých vazebných skupin k chromozómům, jejich orientace a rovněž poloha čtyř malých LG byla zjištěna technikou Fluorescenční In Situ Hybridizace spojené s tyamidovou amplifikací (FISH-TSA) (Krylov et al. 2007). Jednotlivé markery byly vyhledány ve scaffoldech v databázích Ensembl (www.ensembl.org/Xenopus_tropicalis) a GenBank (<http://www.ncbi.nlm.nih.gov>). Celkem byla zjištěna poloha 69 cDNA markerů o minimální délce 900bp. Pozice jednotlivých sond na chromozómech byla vyjádřena jako poměr jeho vzdálenosti od centromery a celkovou délkou příslušného raménka (RCD, Relative Distance From Centromere). Pro každý marker byla zjištěna poloha v alespoň šesti různých mitózách.

Pro každou vazebnou skupinu bylo náhodně vybráno několik markerů, které vždy mapovaly do jednoho chromozómu. Jejich pořadí rovněž umožnilo zjistit orientaci jednotlivých LG (p-q či q-p).

Malé vazebné skupiny byly metodou FISH-TSA přiřazeny k chromozómům 3, 5 a dvě do chromozómu 8. Porovnáním vazebné a fyzické mapy bylo zjištěno, že jsou všechny chromozómy pokryty vazebnými markery po celé své délce. Jedinou výjimkou je krátké raménko chr.2, který neobsahuje žádný vazebný marker a první marker příslušné LG se nachází na centromere. Krátké raménko chr.2 tedy není součástí vazebné mapy. Pravděpodobným důvodem nemožnosti zahrnout tuto oblast do vazebné mapy mohl být nízký až nulový obsah polymorfních SSLP markerů, a s tím i změněná rekombinační frekvence v této oblasti. V době vydání výše uvedeného článku existovalo velké podezření, že se v této oblasti nachází gen určující pohlaví. Studie (Mácha et al. 2012) navíc poukázala na velkou syntenii chr. 2 u *X. tropicalis* s lidským chromozómem X. Poloha sex určující oblasti byla později určena na krátkém raménku chromozómu 7 (Olmstead et al. 2010), kde se ve vazebné mapě nacházela poměrně velká (15cM) oblast bez jediného markeru.

Vazebná skupina	1	2	3	4	5	6	7	8	9	10
Orientace LG	q-p	p-q	p-q	p-q	q-p	q-p	p-q	p-q	q-p	q-p
Chromozóm (Tymowska 1973)	1	2	9	3	4	8	5	10	6	7
Chromozóm (Khokha et al. 2009)	1	2	3	4	5	6	7	8	9	10

Tabulka 1: přiřazení jednotlivých vazebných skupin k chromozómům

Zjištěná poloha některých markerů neodpovídala assembly v.4.1 a tím vypovídala o její možné chybnosti. U celkem 45 scaffoldů byla zjištěna poloha několika markerů ve dvou a více vazebných skupinách. 27 scaffoldů bylo v té době nově vznikající assembly v.5.0 rozděleno na více částí. U zbylých scaffoldů nebyl jejich chimerismus potvrzen.

Vytvořená vazebná mapa se stala důležitým nástrojem pro genetické i genomické studie u *Xenopus tropicalis*, potvrdila assembly v.4.1 (Hellsten et al. 2010) a rovněž sloužila jako důležitý zdroj informací pro nově vzniklou assembly v.7.1.

4.2 Příprava malovacích sond druhu *X. tropicalis* z mikrodisektovaných chromozómů a rekonstrukce tetraploidního karyotypu *X. laevis* metodou Zoo-FISH

K oddělení příbuzných druhů *Xenopus tropicalis* a *Xenopus laevis* došlo přibližně před padesáti až šedesáti milióny let (Evans et al. 2004). I přes tento velmi dlouhý časový úsek je při použití mezidruhově hybridizační sondy (Zoo-FISH) stále možné zjistit velkou podobnost mezi jejich chromozómy a rovněž je možné potvrdit tetraploidní původ druhu *Xenopus laevis*.

Pro studii příbuznosti metodou Zoo-FISH byly z primární testikulární buněčné kultury *X. tropicalis* získány metafázní chromozómy. Fixované buňky byly nakapány na polyethylen-naftalovou membránu, nabarveny Giemsou a mikrodisektovány PALM MicroLaser systémem, který byl spojen s invertovaným mikroskopem. Od každého chromozómu bylo vyříznuto přibližně 15 kopií.

Malovací sondy byly připraveny dvěma odlišnými způsoby. Prvním byla celogenomová amplifikace za použití GenomePlex Single cell WGA (WGA4) kitu s následným značením 20ng amplifikačního produktu Dig-11-dUTP za pomoci GenomePlex DNA Reamplifikačního Kitu (WGA3). Druhým způsobem byly dvě po sobě jdoucí DOP-PCR s Dig-11-dUTP (Kubickova et al. 2002). Tento postup se nakonec ukázal jako neúspěšný. Následná hybridizace testovací sondy z chromozómu *X. tropicalis* (Xt) 9 na Xt kulturu vedla k označení pouze malé oblasti, a to v sekundárním přerušení příslušného chromozómu. Při použití sond vyrobených prvním postupem došlo naopak k označení téměř celých příslušných chromozómů, nicméně ani v tomto případě nebyl nabarvený celý chromozóm. Důvodem mohl být zvýšený obsah repetitivních sekvencí na chromozómu 9, které byly následně inhibovány kompetitorovou DNA.

Sondy z chromozómů Xt 1, Xt 3, Xt 5, Xt 6, Xt 7, Xt 8 a Xt 10 vždy značily vždy pouze příslušný Xt chromozóm. Sondy z chromozómů Xt 2 a Xt 4 bohužel nespecificky značily i další chromozomální oblasti a nebyly z tohoto důvodu dále zahrnuty do studie. Pravděpodobným důvodem bylo přimíchání nesprávného chromozómu do mikrodisekce, což bylo zapříčiněno obtížným rozpoznáním těchto chromozómů. Pro možnou mikrodisekci těchto dvou chromozómů bude nezbytně nutné jejich předchozí bandování.

Po otestování specifity byly jednotlivé Xt sondy hybridizovány s chromozomy *Xenopus laevis* (Xl) získaných ze splenocytů. Xt 1, Xt 5, Xt 6 a Xt 8 značily chromozomální kvartety druhu *Xenopus laevis* (viz tabulka). Nejkomplikovanější situace nastala v případě barvení chromozómem Xt 7, který značil celý kvartet Xl 7a Xl 10, dále satelitní oblast Xl 12 a krom toho byly označeny menší oblasti na mnoha dalších chromozómech.

<i>Xenopus tropicalis</i> (Xt)	1	3	5	6	7	8	9	10
<i>Xenopus laevis</i> (Xl)	chr1+ chr 2	qchr 12+ chr 16	chr 13+ chr 17	chr 6+ chr 9	chr 7+ chr 10+ satelit chr 12+ p13	chr 11+ chr 14	1/2q chr 15+ 1/2q chr 18	Pericentri- cké oblasti chr.14+chr 18

Tabulka 2: Přiřazení příslušných chromozómů podle Zoo-FISH. (Xt) *Xenopus tropicalis*, (Xl) *Xenopus laevis*, (1-10,1-18) označení chromozómu

Poměrně malá označená oblast na Xl chromozómech při použití sondy Xt9 může být způsobena stejným důvodem, jako při značení Xt chromozómů- tedy vysokým obsahem repetice a jejich inhibicí kompetitorem. Možným důvodem může být i to, že chromozómy u *X. laevis* postrádají sekundární konstrikcii vyskytující se na Xt9.

Naše výsledky získané ze Zoo-FISH pokusů podporují pravdivost teorie vzniku *Xenopus laevis* ze dvou diploidních druhů ($2n=20$), u kterých došlo předešlými nerekipročními přestavbami k translokaci jednoho z chromozómů na jiný chromozómy (Schmid and Steinlein 1991). Vzniklé druhy tak měly 18 ($2n$) chromozómů, a jejich křížení dalo vzniknout novému druhu s 36 chromozómy. S největší pravděpodobností podstoupil tyto přestavby Xt 10, který se nevyskytuje u žádného z příbuzných druhů. Sonda Xt10 značila pericentrické oblasti Xl 14+Xl 18, což svědčí o přestavbách proběhlých na těchto chromozómech. Tyto chromozómy jsou rovněž značeny sondami z Xt 8 a Xt 9, pravděpodobně tedy došlo v evoluci k translokaci Xt10 na tyto dva chromozómy.

Dvě sondy (Xt 1 a Xt 6) barvicí chromozomální kvartety odpovídaly studii o rekonstrukci tetraploidního karyotypu u *Xenopus laevis* s replikačním pruhováním (Schmid and Steinlein 1991). Bohužel se nepodařilo získat sondy z chromozómů Xt 2 a Xt 4, které by na základě dalších výzkumů (Tlapakova et al. 2005, Krylov et al. 2007) značily Xl 3+Xl 8, respektive Xl 4+Xl 5.

Tímto výzkumem byla zavedena mikrodisekce chromozómů u druhu *Xenopus tropicalis* a rovněž byl nalezen nejvhodnější typ celogenomové amplifikace pro následné použití jak při výrobě paintovacích sond, tak i pro další molekulárně biologické postupy.

4.3 Efektivní vysokorychlostní sekvenace laserově mikrodisektovaného chromozomálního ramene

Genová assembly představuje u každého modelového organismu důležitý nástroj pro zjišťování genových funkcí či pro evoluční studie. U modelového organismu *Xenopus tropicalis* existují v současné době dvě verze assembly (v.4.1 a v.7.1), ovšem ani jedna z těchto dvou verzí plně nepokrývá jeho deset chromozómů. Hlavním důvodem je velké množství repetice v genomu a tudíž i neschopnost assemblačních programů spojit jednotlivé scaffolds do větších celků. V novější verzi assembly (v.7.1) byly spojeny současné poznatky z nové genetické mapy (Wells et al. 2011) a syntenie s kuřetem, a vzniklá assembly obsahuje 10 velkých scaffoldů odpovídajících deseti chromozómům. I v této verzi je ovšem velké množství scaffoldů nepřirazených- v anglické terminologii se jim říká „orphan scaffolds“. Dalším důvodem nemožnosti dokončit assembly je i nedokonalost vazebné mapy. Ta obsahuje dvě velké oblasti zcela bez markerů- celé dlouhé raménko chromozómu 2 a 15 cM na distálním konci krátkého raménka chromozómu 7. Již pouhým porovnáním v.7.1 a fyzické mapy krátkého raménka chr.2 je možné zjistit, že jsou sice jednotlivé geny přiřazeny správně do chromozómu (pravděpodobně na základě syntenie), nicméně jejich pořadí si neodpovídá. Je tedy zřejmé, že pro dokončení genetické mapy a i pro dokončení assembly je nutné použít další, novější postupy.

V naší studii jsme se věnovali krátkému raménku chromozómu 7, kde se pravděpodobně nachází gen určující pohlaví (Olmstead et al. 2010). Tento chromozóm je snadno rozpoznatelný pod mikroskopem, a to díky sekundární restrikci nacházející se blízko centromery na jeho dlouhém raménku. Ve spolupráci s Veterinárním ústavem v Brně bylo mikrodisektováno 15 kopií krátkého raménka chromozómu 7 za použití PALM MicroLaser systému (P.A.L.M. GmbH, Bernried, Germany), který byl spojen s invertovaným mikroskopem. Z důvodu možné ztráty části chromozómu při vyřezávání laserovým paprskem byl řez veden nad sekundární restrikcí na dlouhém raménku chr.7. Jako zdroj chromozómů sloužila primární testikulární buněčná kultura *X. tropicalis*.

Vzhledem k velmi malému množství získaného materiálu byla získaná DNA amplifikována dvěma amplifikačními kity- WGA4 GenomePlex Single Cell Kit (Sigma-Aldrich) a následně WGA3 (GenomePlex WGA Reamplification Kit, Sigma-Aldrich). Tento systém se nám osvědčil v předchozí studii (Krylov et al. 2010). Průměrná délka amplifikovaných fragmentů činila 500-600bp.

Získaná DNA byla sekvenována NGS systémem Illumina, a to ve spolupráci s MRC National Institute for Medical Research v Londýně. Výsledkem bylo získáno přibližně 35

miliónů přečtených sekvencí o délce 80 bází. Ty byly bioinformačními metodami namapovány algoritmem Bowtie s maximálně dvěma odchylkami v sekvenci do obou verzí assembly (v.4.1 a v.7.1). Poloha zásahů do genomu byla zobrazena pomocí programu GBrowse. Získané sekvence byly rovněž srovnány se známými mRNA *X. tropicalis*.

Do starší verze assembly (v.4.1) se podařilo unikátně namapovat 3.900.340 Illumina čtení. Pro každý scaffold ležící na genetické mapě chromozómu 7 byl spočítán poměr množství zásahů vzhledem k jeho délce. Všechny scaffoldy namapované fyzicky metodou FISH-TSA do p raménka chr.7 měly tento poměr vysoký (více než 29 zásahů/kb), stejně jako většina scaffoldů ve vazebné mapě s příslušnou polohou. Tento poměr rapidně klesl v oblasti centromery (69cM). Fyzicky namapovaný scaffold 297, ležící přibližně v oblasti řezu těsně pod centromerou se 17 zásahy/kb byl určen jako hraniční. Všechny scaffoldy s menším počtem zásahu byly brány jako falešně pozitivní. Jednalo se o 6,3% z celkového počtu zásahů.

67% zásahů mapovalo do scaffoldů nacházejících se v genetické mapě na p ramenu chr.7, 3,5% do scaffoldů namapovaných do jiných oblastí než 7p a 22,5% do scaffoldů bez známé polohy. Předpokládáme, že vzhledem k vysokému poměru počtu zásahů na kb jsou scaffoldy z obou těchto skupin rovněž umístěny v mikrodisektované oblasti chromozómu 7.

Poloha veškerých zásahů v jednotlivých scaffoldech byla zobrazena pomocí programu GBrowse. Přibližně 15% ze všech scaffoldů s polohou na 7p bylo pokryto zásahy nerovnoměrně a vyskytovaly se v nich velké oblasti (přes 100kb) bez jediného zásahu. Vzhledem k tomu, že se v těchto nepokrytých oblastech nacházejí geny a že se tedy nejedná o velké repetitivní oblasti, kam by nemohly zásahy mapovat unikátně, předpokládáme, že jsou tyto scaffoldy hybridní, tj. že se jejich nepokryté části nacházejí v jiných oblastech, než 7p. Tato domněnka byla potvrzena i fyzickým mapováním genů ze scaffoldů 75, 266 a 270. Z každého scaffoldu byl vybrán jeden gen z oblasti s velkým množstvím zásahů a jeden z oblasti bez zásahů. Zatímco geny z pokryté oblasti vždy mapovaly do krátkého raménka chr.7, ty z části bez zásahů mapovaly do jiných chromozómů.

V případě assembly 7.1 mapovalo do superscaffoldu 7 přibližně 80% zásahů, a to většinou do oblasti 0-60Mb odpovídající krátkému raménku. V 7p se nacházelo několik oblastí zcela bez zásahů a naopak v ostatních velkých chromozómech a i 7q se vyskytovaly oblasti s velkým pokrytím, svědčících o špatně sestavené assembly v7.1. Rovněž se nám podařilo přiřadit 14 menších scaffoldů bez známé polohy ke krátkému raménku chr.7.

Touto studií byla v naší laboratoři vyvinuta nová metoda pro sekvenování a mapování mikrodisektovaných částí chromozómů. V porovnání s dřívější podobnou studií, kde byly mikrodisektovaná chromozomální raménka amplifikována pomocí DOP-PCR a sekvenována

454- NGS technologií bylo získáno 30.000x víc čtení a 10.000x více jich unikátně mapovalo do genomu. Podobných výsledků jako našich bylo dosaženo doposud pouze při sekvenování chromozómů získaných pomocí průtokové cytometrie (Hernandez et al. 2012).

Metodu sekvenování mikrodisektovaných chromozómů a jejich částí lze zcela jistě použít nejen pro ověřování genetické mapy či assembly, ale i pro zjišťování polohy genů v chromozómech u organismů s neosekvenovaným genomem. Další aplikace techniky by mohla být v libovolném výzkum vyžadující molekulární analýzu z velmi malého množství buněk, jako je klinická cytogenetika, onkologie a prenatalní diagnostika. Hlavní výhodou tohoto systému je možnost rozpoznání mikrodisektovaného chromozómu pod mikroskopem, která je v porovnání s průtokovou cytometrií výrazně snadnější.

4.4 Velké pseudoautosomální oblasti na pohlavních chromozómech žáby *Silurana (Xenopus) tropicalis*

Xenopus tropicalis má homomorfní, tedy nerozpoznatelné pohlavní chromozómy. U velkého množství druhů došlo u pohlavních chromozómů k jejich rozlišení, a to snížením rekombinace mezi jejími nehomologními částmi s následným hromaděním mutací, inverzemi, delecemi či naopak insercemi DNA v jejich blízkosti (Charlesworth 1991). Právě postupným hromaděním rozdílů mezi chromozómy a s tím spojeným zvětšováním nehomologních oblastí mohou v evoluci vzniknout odlišné chromozómy. U druhu *X. tropicalis* se tak nestalo. Pravděpodobnou příčinou byla mnohonásobná, více než 32 násobná, změna pohlavních chromozómů (Schmid et al. 2010, Evans and Charlesworth 2013). Dalším důvodem mohla být opakovaná rekombinace mezi pohlavními chromozómy po celé jejich délce. Takové chromozómy obsahují velké pseudoautosomální oblasti, které homologní rekombinaci nezabraňují (Stöck et al. 2011).

Přestože má *X. tropicalis* osekvenovaný genom (Hellsten et al. 2010), nebyla u něj po sekvenaci charakterizována oblast určující pohlaví. (Olmstead et al. 2010) našli 22 markerů se silnou vazbou na pohlaví a podle vazebné mapy (Wells et al. 2011) leží některé z těchto markerů na distálním raménku p ramena chromozómu 7, v blízkosti 15cM oblasti, která neobsahuje jediný vazebně mapovaný marker.

Cílem naší práce bylo blíže charakterizovat oblast se silnou vazbou k pohlaví. K tomu byla použita NGS technologie Illumina vyhledávající SNP markery, tzv. „RAD tags“, a to celkem u čtyř samců a čtyřech samic. Základním předpokladem studie bylo, že do genomu bude mapovat větší množství sekvencí získaných sekvenací ze samičí DNA, protože současná

assembly byla sestavena genetického materiálu pocházejícího z tohoto pohlaví. Navíc je samice heterogametním pohlavím, samec by tedy měl některé ze sekvencí postrádat. Tento předpoklad se bohužel nepodařilo prokázat, v obou případech byl do genomu namapován přibližně stejný poměr NGS čtení (87,8-90,8% u samic oproti 85-90,8% u samců). Absolutní počet namapovaných čtení se sice u odlišných pohlaví lišil, ovšem průměrně bylo namapováno u samic o 1,1-3,3% víc čtení do každého z deseti chromozómů, s 2,5% u chromozómu 7, což nesvědčilo o větší odchylce u pohlavního chromozómu.

Ani NGS sekvenování krátkého raménka chromozómu 7 nepřineslo signifikantní výsledky. Mikrodisektované chromozómy byly samčí a toto pohlaví by mělo scaffoldy se silnou vazbou na pohlaví postrádat. To se ovšem prokázalo pouze u dvou ze dvanácti scaffoldů ze studie (Olmstead et al. 2010). Jednalo se o scaffoldy 662 a 278 z assembly v.7.1. V jednom z těchto scaffoldů by se tedy mohl nacházet gen určující pohlaví. Zvláštní ovšem je, že se oproti našemu předpokladu na samčích chromozómech vyskytovaly všechny ostatní scaffoldy se silnou vazbou na pohlaví.

Mezi samci a samicemi nebyly nalezeny významné rozdíly ani v sekvencích jednotlivých markerů použitých ve studii (Olmstead et al. 2010), a žádná z frekvenčních odlišností nebyla fixována u samčího pohlaví. Na druhou stranu se v současných verzích assembly nachází pouze asi 80% ze všech bází a sekvenování mikrodisektovaného raménka krátkého raménka chr 7. prokázala, že se ve verzi v.7.1 nachází na krátkém raménku chromozómu 7 několik oblastí, které jsou sestaveny špatně, nebo vůbec. Je tedy možné, že se gen určující pohlaví nachází v částech genomu, o kterých nemáme žádné informace.

Celkově naše výsledky nasvědčují, že je oblast určující pohlaví na p raménku chr.7 velmi malá a rozdíly pohlavních chromozómů jsou minimální. Tento výsledek je shodný se studií (Uno et al. 2008), ve které se nepodařilo bandovacími technikami nalézt rozdíly mezi samčími a samičími chromozómy u *X. tropicalis*. Přesná poloha genu určujícího pohlaví zůstává i nadále nejasná. Navíc není známý ani přesný způsob, jakým je pohlaví určeno. Např. u druhu *Danio rerio* se ukazuje, že má pohlaví určené sice geneticky, ovšem větším množstvím genů nacházejících se na různých chromozómech (Anderson et al., 2012). O obdobnou komplikovanou situaci by se mohlo jednat i u druhu *X. tropicalis*.

5. Závěr

Veškeré studie zahrnuté v této dizertační práci se týkaly genetiky u *Xenopus tropicalis*. V první studii byla sestavená genetická mapa, která se stala velmi důležitým zdrojem informací pro veškeré následující genetické, genomické, ale i cytologické a evoluční výzkumy. Mapu byla tvořena ze dvou celků- vazebné mapy, která obsahovala 2886 vazebně mapovaných markerů, a mapy fyzické. Ta byla vytvořena v naší laboratoři a její součástí bylo 69 genů s fyzicky zjištěnou polohou. Propojením obou přístupů se podařilo přiřadit všech deset velkých a čtyři malé vazebné skupiny k jednotlivým chromozómům a zjistit jejich orientaci. Přestože se jedná o nejlepší doposud vytvořenou mapu u tohoto druhu, její součástí není krátké raménko chromozómu 2 a rovněž obsahovala mezeru na 15 distálních centimorganech krátkého raménka chromozómu 7. Toto místo vykazovalo ve studii (Olmstead et al. 2010) velmi silnou vazbu na pohlaví a jeho vazebné mapování je velmi obtížné.

V rámci další studie se nám podařilo vyvinout novou metodu, kterou se nám podařilo ke krátkému raménku chromozómu 7 přiřadit přes 200 nových scaffoldů z assembly v.4.1, nalézt chybně přiřazené scaffoldy a rovněž poukázat na chybnost sestavení u přibližně 15% scaffoldů v této verzi genomu. Chyby byly touto metodou nalezeny i u novější verze (v.7.1). Základním principem nově vyvinutého přístupu je mikrodisekce přibližně patnácti zvolených chromozomálních oblastí a následující dvojité celogenomové amplifikace získané DNA kity WGA4 a WGA3. Získaný materiál je osekvenován metodou Illumina, jednotlivá čtení jsou bioinformatickými přístupy mapována do genomu a zásahy do něj jsou graficky zobrazeny. U špatně sestavených scaffoldů je pokryta zásahy pouze jeho část, u správně sestavených jsou zásahy rozmístěny rovnoměrně. Přístup spojující mikrodisekci, amplifikaci a sekvenaci byla pro genetické mapování použita poprvé a i samotná metodika byla oproti dříve používaným postupům značně vylepšena. Počet jednotlivých čtení až 3000x převyšoval počet čtení získaných v podobné studii u myši. Je zřejmé, že je možné tuto metodu použít u libovolného druhu, jediným limitujícím faktorem je možnost rozeznat jednotlivé chromozómy pod mikroskopem.

Při sestavování výše uvedeného protokolu jsme vycházeli z naší předchozí studie srovnávající cytologickými metodami chromozómy druhů *X. tropicalis* a *X. laevis*. V této studii byly mikrodisektovány chromozómy z buněčné linie druhu *X. tropicalis*. Získaná DNA byla amplifikována dvěma metodami- dvěma po sobě jdoucími DOP-PCR a celogenomovou amplifikací kity WGA4+WGA3. Druhá amplifikace u obou metod zároveň sloužila k označení DNA pomocí Dig-11-dUTP. Z výše uvedených postupů se první (DOP-PCR) ukázal jako zcela nevhodný, protože sonda značila příslušný chromozóm nedostatečně. Druhou technikou se nám

podařilo vyrobit sondy pro všechny chromozómy *X. tropicalis*, s výjimkou chr.2 a 4. V jejich případě došlo k chybě při mikrodisekci, při které byly vyřezávané chromozómy namíchány s jinými. Sondy z ostatních chromozómů byly použity k značení chromozómů *X. laevis*. Touto studií jsme zjistili, že jsou chromozómy *X. tropicalis* č. 1, 5, 6 a 8 sekvenčně velmi podobné s příslušnými kvartety u *X. laevis*. Rovněž jsme poukázali na fakt, že v evoluci došlo ke vzniku dvou druhů po translokaci chromozómu Xt 10 na dva jiné chromozómy (pravděpodobně Xt8 a Xt9). Tyto druhy se křížily za vzniku nového druhu s 36 chromozómy.

I poslední studie byla evolučního charakteru a týkala se pohlavních chromozómů. Pomocí metody RAD tags a rovněž dat získaných ze sekvenace mikrodisektovaných p ramének chromozómu 7 jsme zjistili, že pohlavní rozdíly jsou u tohoto chromozómu minimální. Pravděpodobným důvodem je mnohonásobná translokace genu určujícího pohlaví, ke které u obojživelníků často dochází. Crossing-over probíhá prakticky po celé jejich délce a proto jsou pohlavní rozdíly mezi nimi minimální. Jediné větší rozdíly byly nalezeny u scaffoldů 278 a 662 z assembly v.7.1, které se nenacházely na mikrodisektovaných krátkých raménkách chr.7 ze samčích buněk. Protože data z assembly pocházely ze samice, která je heterogametním pohlavím, je možné, že se na jednom z těchto scaffoldů nachází gen určující pohlaví.

Veškeré nově vynalezené techniky z výše uvedených studií je možné použít i u dalších modelových organismů. Jediným předpokladem je rozpoznatelnost jednotlivých chromozómů pod mikroskopem. Protože je genom *X. tropicalis* sestaven pouze částečně a obsahuje velké množství repetitivních sekvencí, lze předpokládat, že u jiných modelových organismů (myš, kuře) by počet výsledných zásahů mohl být vyšší, než v případě tohoto obojživelníka. Značnou výhodou metody je potřeba pouze malého množství vstupního materiálu pro mikrodisekci, která umožňuje její využití i pro klinické studie v onkologii či prenatální diagnostice.

6. Použitá literatura

- Abu-Daya, A., A. K. Sater, D. E. Wells, T. J. Mohun, and L. B. Zimmerman. 2009. Absence of heartbeat in the *Xenopus tropicalis* mutation muzak is caused by a nonsense mutation in cardiac myosin myh6. *Dev Biol* **336**:20-29.
- Aparicio, S., J. Chapman, E. Stupka, N. Putnam, J. M. Chia, P. Dehal, A. Christoffels, S. Rash, S. Hoon, A. Smit, M. D. Gelpke, J. Roach, T. Oh, I. Y. Ho, M. Wong, C. Detter, F. Verhoef, P. Predki, A. Tay, S. Lucas, P. Richardson, S. F. Smith, M. S. Clark, Y. J. Edwards, N. Doggett, A. Zharkikh, S. V. Tavtigian, D. Pruss, M. Barnstead, C. Evans, H. Baden, J. Powell, G. Glusman, L. Rowen, L. Hood, Y. H. Tan, G. Elgar, T. Hawkins, B. Venkatesh, D. Rokhsar, and S. Brenner. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**:1301-1310.
- Bartholdi, M. F. 1990. Flow cytogenetics. *Pathobiology* **58**:118-128.
- Batzoglou, S., D. B. Jaffe, K. Stanley, J. Butler, S. Gnerre, E. Mauceli, B. Berger, J. P. Mesirov, and E. S. Lander. 2002. ARACHNE: a whole-genome shotgun assembler. *Genome Res* **12**:177-189.
- Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. Rasoloniato, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, M. Chiara E Catenazzi, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoschler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. Huw Jones, G. D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, D. Chris Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Racz, V. H. Rae, S. R. Rawlings, A. Chiva Rodriguez, P. M. Roe, J. Rogers, M. C. Robert Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. Ernest Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. Vandevondele, Y. Verhovsky, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin, and A. J. Smith. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**:53-59.
- Berns, M. W., and A. D. Floyd. 1971. Chromosomal microdissection by laser. A cytochemical and functional analysis. *Exp Cell Res* **67**:305-310.
- Bewick, A. J., D. W. Anderson, and B. J. Evans. 2011. Evolution of the closely related, sex-related genes DM-W and DMRT1 in African clawed frogs (*Xenopus*). *Evolution* **65**:698-712.
- Boland, E. J., A. Pillai, M. W. Odom, and P. Jagadeeswaran. 1994. Automation of the Maxam-Gilbert chemical sequencing reactions. *Biotechniques* **16**:1088-1092, 1094-1085.
- Bonnette, M. D., V. R. Pavlova, D. N. Rodier, L. P. Thompson, E. L. Boone, K. L. Brown, K. M. Meyer, M. B. Trevino, J. R. Champagne, and T. D. Cruz. 2009. dcDegenerate oligonucleotide primed-PCR for multilocus, genome-wide analysis from limited quantities of DNA. *Diagn Mol Pathol* **18**:165-175.

- Brenner, S., M. Johnson, J. Bridgham, G. Golda, D. H. Lloyd, D. Johnson, S. Luo, S. McCurdy, M. Foy, M. Ewan, R. Roth, D. George, S. Eletr, G. Albrecht, E. Vermaas, S. R. Williams, K. Moon, T. Burcham, M. Pallas, R. B. DuBridge, J. Kirchner, K. Fearon, J. Mao, and K. Corcoran. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* **18**:630-634.
- Carrano, A. V., B. H. Mayall, J. R. Testa, L. K. Ashworth, and J. D. Rowley. 1979. Chromosomal DNA cytophotometry in 20q- nonspecific myeloid disorders. *Cancer Res* **39**:2984-2987.
- Charlesworth, B. 1991. Evolution. When to be diploid. *Nature* **351**:273-274.
- Dayarian, A., T. P. Michael, and A. M. Sengupta. 2010. SOPRA: Scaffolding algorithm for paired reads via statistical optimization. *BMC Bioinformatics* **11**:345.
- de Sá, R. O., and D. M. Hillis. 1990. Phylogenetic relationships of the pipid frogs *Xenopus* and *Silurana*: an integration of ribosomal DNA and morphology. *Mol Biol Evol* **7**:365-376.
- Dietmaier, W., A. Hartmann, S. Wallinger, E. Heinmöller, T. Kerner, E. Endl, K. W. Jauch, F. Hofstädter, and J. Rüschoff. 1999. Multiple mutation analyses in single tumor cells with improved whole genome amplification. *Am J Pathol* **154**:83-95.
- Doležel, J., J. Vrána, J. Safář, J. Bartoš, M. Kubaláková, and H. Simková. 2012. Chromosomes in the flow to simplify genome analysis. *Funct Integr Genomics* **12**:397-416.
- Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korch, and S. Turner. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* **323**:133-138.
- Emmert-Buck, M. R., R. F. Bonner, P. D. Smith, R. F. Chuaqui, Z. Zhuang, S. R. Goldstein, R. A. Weiss, and L. A. Liotta. 1996. Laser capture microdissection. *Science* **274**:998-1001.
- Evans, B. J., and B. Charlesworth. 2013. The effect of nonindependent mate pairing on the effective population size. *Genetics* **193**:545-556.
- Evans, B. J., D. B. Kelley, R. C. Tinsley, D. J. Melnick, and D. C. Cannatella. 2004. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Mol Phylogenet Evol* **33**:197-213.
- Froehlich, E., J. S. Mandeville, C. M. Weinert, L. Kreplak, and H. A. Tajmir-Riahi. 2011. Bundling and aggregation of DNA by cationic dendrimers. *Biomacromolecules* **12**:511-517.
- Gall, J. G., and M. L. Pardue. 1969. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci U S A* **63**:378-383.
- Gao, S., W. K. Sung, and N. Nagarajan. 2011. Opera: reconstructing optimal genomic scaffolds with high-throughput paired-end sequences. *J Comput Biol* **18**:1681-1691.
- Gharizadeh, B., Z. S. Herman, R. G. Eason, O. Jejelowo, and N. Pourmand. 2006. Large-scale pyrosequencing of synthetic DNA: a comparison with results from Sanger dideoxy sequencing. *Electrophoresis* **27**:3042-3047.
- Graf, J. D. 1989. Genetic mapping in *Xenopus laevis*: eight linkage groups established. *Genetics* **123**:389-398.
- Graf, J. D., and H. R. Kobel. 1991. Genetics of *Xenopus laevis*. *Methods Cell Biol* **36**:19-34.
- Grey, D., J. M. Hamilton-Miller, and W. Brumfitt. 1979. Incidence and mechanisms of resistance to trimethoprim in clinically isolated gram-negative bacteria. *Chemotherapy* **25**:147-156.
- Griffin, B. E., M. Fried, and A. Cowie. 1974. Polyoma DNA: a physical map. *Proc Natl Acad Sci U S A* **71**:2077-2081.
- Hellsten, U., R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N. H. Putnam, S. Shu, L. Taher, I. L. Blitz, B. Blumberg, D. S. Dichmann, I. Dubchak, E. Amaya, J. C. Detter, R. Fletcher, D. S. Gerhard, D. Goodstein, T. Graves, I. V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S. M. Lucas, P. E. Mead, T. Mitros, H. Ogino, Y. Ohta, A. V. Poliakov, N. Pollet, J. Robert, A. Salamov, A. K. Sater, J. Schmutz, A. Terry, P. D. Vize, W. C. Warren, D. Wells,

- A. Wills, R. K. Wilson, L. B. Zimmerman, A. M. Zorn, R. Grainger, T. Grammer, M. K. Khokha, P. M. Richardson, and D. S. Rokhsar. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**:633-636.
- Hernandez, P., M. Martis, G. Dorado, M. Pfeifer, S. Gálvez, S. Schaaf, N. Jouve, H. Šimková, M. Valárik, J. Doležal, and K. F. Mayer. 2012. Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J* **69**:377-386.
- Hirsch, N., L. B. Zimmerman, J. Gray, J. Chae, K. L. Curran, M. Fisher, H. Ogino, and R. M. Grainger. 2002. *Xenopus tropicalis* transgenic lines and their use in the study of embryonic induction. *Dev Dyn* **225**:522-535.
- Huse, S. M., J. A. Huber, H. G. Morrison, M. L. Sogin, and D. M. Welch. 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* **8**:R143.
- Höckner, M., M. Erdel, A. Spreiz, G. Utermann, and D. Kotzot. 2009. Whole genome amplification from microdissected chromosomes. *Cytogenet Genome Res* **125**:98-102.
- Jiang, J., and B. S. Gill. 2006. Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. *Genome* **49**:1057-1068.
- Jiang, Z., X. Zhang, R. Deka, and L. Jin. 2005. Genome amplification of single sperm using multiple displacement amplification. *Nucleic Acids Res* **33**:e91.
- Khokha, M. K., V. Krylov, M. J. Reilly, J. G. Gall, D. Bhattacharya, C. Y. Cheung, S. Kaufman, D. K. Lam, J. Macha, C. Ngo, N. Prakash, P. Schmidt, T. Tlapakova, T. Trivedi, L. Tumova, A. Abu-Day, T. Geach, E. Vendrell, H. Ironfield, L. Sinzelle, A. K. Sater, D. E. Wells, R. M. Harland, and L. B. Zimmerman. 2009. Rapid gynogenetic mapping of *Xenopus tropicalis* mutations to chromosomes. *Dev Dyn* **238**:1398-1346.
- Khokha, M. K., and G. G. Loots. 2005. Strategies for characterising cis-regulatory elements in *Xenopus*. *Brief Funct Genomic Proteomic* **4**:58-68.
- Kittler, R., M. Stoneking, and M. Kayser. 2002. A whole genome amplification method to generate long fragments from low quantities of genomic DNA. *Anal Biochem* **300**:237-244.
- Kochan, K. J., D. A. Wright, L. J. Schroeder, J. Shen, and D. C. Morizot. 2003. Genetic linkage maps of the West African clawed frog *Xenopus tropicalis*. *Dev Dyn* **226**:99-102.
- Kondo, M., E. Nagao, H. Mitani, and A. Shima. 2001. Differences in recombination frequencies during female and male meioses of the sex chromosomes of the medaka, *Oryzias latipes*. *Genet Res* **78**:23-30.
- Kooi, M. W., J. A. Aten, J. Stap, J. B. Kipp, and G. W. Barendsen. 1984. Preparation of chromosome suspensions from cells of a solid experimental tumour for measurement by flow cytometry. *Cytometry* **5**:547-549.
- Krylov, V., S. Kubickova, J. Rubes, J. Macha, T. Tlapakova, E. Seifertova, and N. Sebkova. 2010. Preparation of *Xenopus tropicalis* whole chromosome painting probes using laser microdissection and reconstruction of *X. laevis* tetraploid karyotype by Zoo-FISH. *Chromosome Res* **18**:431-439.
- Krylov, V., T. Tlapakova, and J. Macha. 2007. Localization of the single copy gene *Mdh2* on *Xenopus tropicalis* chromosomes by FISH-TSA. *Cytogenet Genome Res* **116**:110-112.
- Kubickova, S., H. Cernohorska, P. Musilova, and J. Rubes. 2002. The use of laser microdissection for the preparation of chromosome-specific painting probes in farm animals. *Chromosome Res* **10**:571-577.
- Kölble, K. 2000. The LEICA microdissection system: design and applications. *J Mol Med (Berl)* **78**:B24-25.
- Ladanyi, A., F. Sipos, D. Szoke, O. Galamb, B. Molnar, and Z. Tulassay. 2006. Laser microdissection in translational and clinical research. *Cytometry A* **69**:947-960.
- Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**:R25.
- Li, H., J. Ruan, and R. Durbin. 2008a. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**:1851-1858.

- Li, R., Y. Li, K. Kristiansen, and J. Wang. 2008b. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**:713-714.
- Lledó, B., J. Ten, D. Rodríguez-Arnedo, J. Llácer, and R. Bernabeu. 2008. Preimplantation genetic diagnosis of X-linked retinoschisis. *Reprod Biomed Online* **16**:886-892.
- Lovmar, L., and A. C. Syvänen. 2006. Multiple displacement amplification to create a long-lasting source of DNA for genetic studies. *Hum Mutat* **27**:603-614.
- Malcov, M., T. Naiman, D. B. Yosef, A. Carmon, N. Mey-Raz, A. Amit, I. Vagman, and Y. Yaron. 2007. Preimplantation genetic diagnosis for fragile X syndrome using multiplex nested PCR. *Reprod Biomed Online* **14**:515-521.
- Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376-380.
- Matsson, P., and B. Rydberg. 1981. Analysis of chromosomes from human peripheral lymphocytes by flow cytometry. *Cytometry* **1**:369-372.
- Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* **74**:560-564.
- Micke, P., A. Ostman, J. Lundeberg, and F. Ponten. 2005. Laser-assisted cell microdissection using the PALM system. *Methods Mol Biol* **293**:151-166.
- Mikamo, K., and E. Witschi. 1963. [BREEDING EXPERIMENTS WITH SEX-CONVERTED CLAWED TOADS (XENOPUS LAEVIS)]. *Experientia* **19**:536-537.
- Mácha, J., R. Teichmanová, A. K. Sater, D. E. Wells, T. Tlapáková, L. B. Zimmerman, and V. Krylov. 2012. Deep ancestry of mammalian X chromosome revealed by comparison with the basal tetrapod *Xenopus tropicalis*. *BMC Genomics* **13**:315.
- Olmstead, A. W., A. Lindberg-Livingston, and S. J. Degitz. 2010. Genotyping sex in the amphibian, *Xenopus (Silurana) tropicalis*, for endocrine disruptor bioassays. *Aquat Toxicol* **98**:60-66.
- Peng, W., H. Takabayashi, and K. Ikawa. 2007. Whole genome amplification from single cells in preimplantation genetic diagnosis and prenatal diagnosis. *Eur J Obstet Gynecol Reprod Biol* **131**:13-20.
- Poinar, G. O., and D. C. Cannatella. 1988. Response:significance of frog in amber. *Science* **239**:1478.
- Pop, M., A. Phillippy, A. L. Delcher, and S. L. Salzberg. 2004. Comparative genome assembly. *Brief Bioinform* **5**:237-248.
- Rens, W., G. A. Boschman, J. M. Hoovers, E. M. Manders, R. M. Slater, J. Stap, and J. A. Aten. 1994. Flow cytometric detection of chromosome abnormalities by measuring centromeric index, DNA content, and DNA base composition. *Anal Cell Pathol* **6**:359-375.
- Sanger, F., and A. R. Coulson. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**:441-448.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**:5463-5467.
- Scacheri, P. C., O. Rozenblatt-Rosen, N. J. Caplen, T. G. Wolfsberg, L. Umayam, J. C. Lee, C. M. Hughes, K. S. Shanmugam, A. Bhattacharjee, M. Meyerson, and F. S. Collins. 2004. Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc Natl Acad Sci U S A* **101**:1892-1897.
- Scherthan, H., T. Cremer, U. Arnason, H. U. Weier, A. Lima-de-Faria, and L. Frönicke. 1994. Comparative chromosome painting discloses homologous segments in distantly related mammals. *Nat Genet* **6**:342-347.

- Schmid, M., and C. Steinlein. 1991. Chromosome banding in Amphibia. XVI. High-resolution replication banding patterns in *Xenopus laevis*. *Chromosoma* **101**:123-132.
- Schmid, M., C. Steinlein, J. P. Bogart, W. Feichtinger, P. León, E. La Marca, L. M. Díaz, A. Sanz, S. H. Chen, and S. B. Hedges. 2010. The chromosomes of terraranan frogs. Insights into vertebrate cytogenetics. *Cytogenet Genome Res* **130-131**:1-568.
- Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. Kent, and L. E. Hood. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* **321**:674-679.
- Smith, W. C., and R. M. Harland. 1991. Injected Xwnt-8 RNA acts early in *Xenopus* embryos to promote formation of a vegetal dorsalizing center. *Cell* **67**:753-765.
- Spits, C., C. Le Caignec, M. De Rycke, L. Van Haute, A. Van Steirteghem, I. Liebaers, and K. Sermon. 2006. Whole-genome multiple displacement amplification from single cells. *Nat Protoc* **1**:1965-1970.
- Stöck, M., A. Horn, C. Grossen, D. Lindtke, R. Sermier, C. Betto-Colliard, C. Dufresnes, E. Bonjour, Z. Dumas, E. Luquet, T. Maddalena, H. C. Sousa, I. Martinez-Solano, and N. Perrin. 2011. Ever-young sex chromosomes in European tree frogs. *PLoS Biol* **9**:e1001062.
- Telenius, H., N. P. Carter, C. E. Bebb, M. Nordenskjöld, B. A. Ponder, and A. Tunnacliffe. 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13**:718-725.
- Thiébaud, C. H., and M. Fischberg. 1977. DNA content in the genus *Xenopus*. *Chromosoma* **59**:253-257.
- Tlapakova, T., V. Krylov, and J. Macha. 2005. Localization, structure and polymorphism of two paralogous *Xenopus laevis* mitochondrial malate dehydrogenase genes. *Chromosome Res* **13**:699-706.
- Tymowska, J. 1973. Karyotype analysis of *Xenopus tropicalis* Gray, Pipidae. *Cytogenet Cell Genet* **12**:297-304.
- Uno, Y., C. Nishida, Y. Oshima, S. Yokoyama, I. Miura, Y. Matsuda, and M. Nakamura. 2008. Comparative chromosome mapping of sex-linked genes and identification of sex chromosomal rearrangements in the Japanese wrinkled frog (*Rana rugosa*, Ranidae) with ZW and XY sex chromosome systems. *Chromosome Res* **16**:637-647.
- van den Engh, G., B. Trask, S. Cram, and M. Bartholdi. 1984. Preparation of chromosome suspensions for flow cytometry. *Cytometry* **5**:108-117.
- Wells, D., T. Escudero, B. Levy, K. Hirschhorn, J. D. Delhanty, and S. Munné. 2002. First clinical application of comparative genomic hybridization and polar body testing for preimplantation genetic diagnosis of aneuploidy. *Fertil Steril* **78**:543-549.
- Wells, D., J. K. Sherlock, A. H. Handyside, and J. D. Delhanty. 1999. Detailed chromosomal and molecular genetic analysis of single cells by whole genome amplification and comparative genomic hybridisation. *Nucleic Acids Res* **27**:1214-1218.
- Wells, D. E., L. Gutierrez, Z. Xu, V. Krylov, J. Macha, K. P. Blankenburg, M. Hitchens, L. J. Bellot, M. Spivey, D. L. Stemple, A. Kowis, Y. Ye, S. Pasternak, J. Owen, T. Tran, R. Slavikova, L. Tumova, T. Tlapakova, E. Seifertova, S. E. Scherer, and A. K. Sater. 2011. A genetic map of *Xenopus tropicalis*. *Dev Biol* **354**:1-8.
- Wetzel, J., C. Kingsford, and M. Pop. 2011. Assessing the benefits of using mate-pairs to resolve repeats in de novo short-read prokaryotic assemblies. *BMC Bioinformatics* **12**:95.
- Yoshimoto, S., E. Okada, T. Oishi, R. Numagami, H. Umemoto, K. Tamura, H. Kanda, T. Shiba, N. Takamatsu, and M. Ito. 2006. Expression and promoter analysis of *Xenopus* DMRT1 and functional characterization of the transactivation property of its protein. *Dev Growth Differ* **48**:597-603.
- Yu, X., K. Guda, J. Willis, M. Veigl, Z. Wang, S. Markowitz, M. D. Adams, and S. Sun. 2012. How do alignment programs perform on sequencing data with varying qualities and from repetitive regions? *BioData Min* **5**:6.
- Zhang, L., X. Cui, K. Schmitt, R. Hubert, W. Navidi, and N. Arnheim. 1992. Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci U S A* **89**:5847-5851.